ELSEVIER

CrossMark

Available online at www.sciencedirect.com

ScienceDirect

Ann. I. H. Poincaré – AN 33 (2016) 1589–1638

ANNALES DE L'INSTITUT HENRI POINCARÉ ANALYSE NON LINÉAIRE

www.elsevier.com/locate/anihpc

KAM for autonomous quasi-linear perturbations of KdV

Pietro Baldi [a], Massimiliano Berti [b,*], Riccardo Montalto [b,c]

[a] *Dipartimento di Matematica e Applicazioni "R. Caccioppoli", Università di Napoli Federico II, Via Cintia, Monte S. Angelo, 80126, Napoli, Italy*

[b] *SISSA, Via Bonomea 265, 34136, Trieste, Italy*

[c] *Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*

Received 16 October 2014; received in revised form 13 July 2015; accepted 22 July 2015

Available online 3 August 2015

Abstract

We prove the existence and the stability of Cantor families of quasi-periodic, small amplitude solutions of *quasi-linear* (i.e. *strongly nonlinear*) autonomous Hamiltonian differentiable perturbations of KdV. This is the first result that extends KAM theory to quasi-linear autonomous and parameter independent PDEs. The core of the proof is to find an approximate inverse of the linearized operators at each approximate solution and to prove that it satisfies tame estimates in Sobolev spaces. A symplectic decoupling procedure reduces the problem to the one of inverting the linearized operator restricted to the normal directions. For this aim we use pseudo-differential operator techniques to transform such linear PDE into an equation with constant coefficients up to smoothing remainders. Then a linear KAM reducibility technique completely diagonalizes such operator. We introduce the "initial conditions" as parameters by performing a "weak" Birkhoff normal form analysis, which is well adapted for quasi-linear perturbations.

MSC: 37K55; 35Q53

Keywords: KdV; KAM for PDEs; Quasi-linear PDEs; Nash–Moser theory; Quasi-periodic solutions

1. Introduction and main results

In this paper we prove the existence and stability of Cantor families of quasi-periodic solutions of Hamiltonian *quasi-linear* (also called "*strongly nonlinear*", e.g. in [25]) perturbations of the KdV equation

$$u_t + u_{xxx} - 6uu_x + \mathcal{N}_4(x, u, u_x, u_{xx}, u_{xxx}) = 0\,, \tag{1.1}$$

under periodic boundary conditions $x \in \mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$, where

$$\mathcal{N}_4(x, u, u_x, u_{xx}, u_{xxx}) := -\partial_x\big[(\partial_u f)(x, u, u_x) - \partial_x((\partial_{u_x} f)(x, u, u_x))\big] \tag{1.2}$$

* Corresponding author.

E-mail addresses: pietro.baldi@unina.it (P. Baldi), berti@sissa.it (M. Berti), riccardo.montalto@sissa.it, riccardo.montalto@math.uzh.ch (R. Montalto).

http://dx.doi.org/10.1016/j.anihpc.2015.07.003

is the most general quasi-linear Hamiltonian (local) nonlinearity. Note that $\mathcal{N}_4$ contains as many derivatives as the linear part ∂_{xxx}. The equation (1.1) is the Hamiltonian PDE $u_t = \partial_x \nabla H(u)$ where ∇H denotes the $L^2(\mathbb{T}_x)$ gradient of the Hamiltonian

$$H(u) = \int_{\mathbb{T}} \frac{u_x^2}{2} + u^3 + f(x, u, u_x)\, dx \tag{1.3}$$

on the real phase space

$$H_0^1(\mathbb{T}_x) := \Big\{ u(x) \in H^1(\mathbb{T}, \mathbb{R}) \ : \ \int_{\mathbb{T}} u(x) dx = 0 \Big\}. \tag{1.4}$$

We assume that the "Hamiltonian density" $f \in C^q(\mathbb{T} \times \mathbb{R} \times \mathbb{R}; \mathbb{R})$ for some q large enough, and that

$$f = f_5(u, u_x) + f_{\geq 6}(x, u, u_x), \tag{1.5}$$

where $f_5(u, u_x)$ denotes the homogeneous component of f of degree 5 and $f_{\geq 6}$ collects all the higher order terms. By (1.5) the nonlinearity $\mathcal{N}_4$ vanishes of order 4 at $u = 0$ and (1.1) may be seen, close to the origin, as a "small" perturbation of the KdV equation

$$u_t + u_{xxx} - 6uu_x = 0, \tag{1.6}$$

which is completely integrable. Actually, the KdV equation (1.6) may be described by global analytic action-angle variables, see [21] and the references therein.

A natural question is to know whether the periodic, quasi-periodic or almost periodic solutions of (1.6) persist under small perturbations. This is the content of KAM theory.

The first KAM results for PDEs have been obtained for 1-d semilinear Schrödinger and wave equations by Kuksin [23], Wayne [33], Craig–Wayne [12], Pöschel [27], see [11,25] and references therein. For PDEs in higher space dimension the theory has been more recently extended by Bourgain [10], Eliasson–Kuksin [13], and Berti–Bolle [6], Geng–Xu–You [14], Procesi–Procesi [30,29], Wang [32].

For *unbounded* perturbations the first KAM results have been proved by Kuksin [24] and Kappeler–Pöschel [21] for KdV (see also Bourgain [9]), and more recently by Liu–Yuan [20], Zhang–Gao–Yuan [34] for derivative NLS, and by Berti–Biasco–Procesi [4,5] for derivative NLW. For a recent survey of known results for KdV, we refer to [15].

The KAM theorems in [24,21] prove the persistence of the finite-gap solutions of the integrable KdV (1.6) under semilinear Hamiltonian perturbations $\varepsilon \partial_x (\partial_u f)(x, u)$, namely when the density f is independent of u_x, so that (1.2) is a differential operator of order 1 (note that in [25] such nonlinearities are called "quasi-linear" and (1.2) "strongly nonlinear"). The key point is that the frequencies of KdV grow as $\sim j^3$ and the difference $|j^3 - i^3| \geq (j^2 + i^2)/2$, $i \neq j$, so that KdV gains (outside the diagonal) two derivatives. This approach also works for Hamiltonian pseudo-differential perturbations of order 2 (in space), using the improved Kuksin's lemma in [20]. However it does *not* work for a general quasi-linear perturbation as in (1.2), which is a nonlinear differential operator of the *same* order (i.e. 3) as the constant coefficient linear operator ∂_{xxx}. Such a strongly nonlinear perturbation term makes the KAM question quite delicate because of the possible phenomenon of formation of singularities in finite time, see Lax [19], Klainerman–Majda [22] for quasi-linear wave equations, see also Section 1.4 of [25]. For example, Kappeler–Pöschel [21] (Remark 3, page 19) wrote: "*It would be interesting to obtain perturbation results which also include terms of higher order, at least in the region where the KdV approximation is valid. However, results of this type are still out of reach, if true at all*".

This paper gives the first positive answer to KAM theory for quasi-linear PDEs, proving the existence of small amplitude, linearly stable, quasi-periodic solutions of (1.1)–(1.2), see Theorem 1.1. As a consequence, for most initial conditions, quasi-linear Hamiltonian perturbations of KdV do not produce formation of singularities in the solutions, and the KAM phenomenon persists! We mention that, concerning the initial value problem for (1.1)–(1.2), there are no results even for the local existence theory. On the other hand, the initial conditions selected by the KAM Theorem 1.1 give rise to global in time solutions. We find it interesting because such PDEs are in general ill-posed in Sobolev spaces.

We also note that (1.1) does not depend on external parameters. Moreover the KdV equation (1.1) is a *completely resonant* PDE, namely the linearized equation at the origin is the linear Airy equation $u_t + u_{xxx} = 0$, which possesses only the 2π-periodic in time solutions

$$u(t,x) = \sum_{j \in \mathbb{Z} \setminus \{0\}} u_j e^{\mathrm{i} j^3 t} e^{\mathrm{i} j x} . \tag{1.7}$$

Thus the existence of quasi-periodic solutions of (1.1) is a purely nonlinear phenomenon (the diophantine frequencies in (1.9) are $O(|\xi|)$-close to integers with $\xi \to 0$) and a perturbation theory is more difficult.

The solutions that we find are localized in Fourier space close to finitely many "tangential sites"

$$S^+ := \{\bar{\jmath}_1, \dots, \bar{\jmath}_\nu\}, \quad S := S^+ \cup (-S^+) = \{\pm j : j \in S^+\}, \quad \bar{\jmath}_i \in \mathbb{N} \setminus \{0\}, \quad \forall i = 1, \dots, \nu . \tag{1.8}$$

The set S is required to be even because the solutions u of (1.1) have to be real valued. Moreover, we also assume the following explicit hypotheses on S:

- ($\mathtt{S}1$) $j_1 + j_2 + j_3 \neq 0$ for all $j_1, j_2, j_3 \in S$.
- ($\mathtt{S}2$) $\nexists j_1, \dots, j_4 \in S$ such that $j_1 + j_2 + j_3 + j_4 \neq 0$, $j_1^3 + j_2^3 + j_3^3 + j_4^3 - (j_1 + j_2 + j_3 + j_4)^3 = 0$.

Theorem 1.1 *(KAM for quasi-linear perturbations of KdV). Given $\nu \in \mathbb{N}$, let $f \in C^q$ (with $q := q(\nu)$ large enough) satisfy (1.5). Then, for all the tangential sites S as in (1.8) satisfying* ($\mathtt{S}1$)–($\mathtt{S}2$)*, the KdV equation (1.1) possesses small amplitude quasi-periodic solutions with diophantine frequency vector $\omega := \omega(\xi) = (\omega_j)_{j \in S^+} \in \mathbb{R}^\nu$, of the form*

$$u(t,x) = \sum_{j \in S^+} 2\sqrt{\xi_j} \cos(\omega_j t + jx) + o(\sqrt{|\xi|}), \quad \omega_j := j^3 - 6\xi_j j^{-1} , \tag{1.9}$$

for a "Cantor-like" set of small amplitudes $\xi \in \mathbb{R}^\nu_+$ with density 1 at $\xi = 0$. The term $o(\sqrt{|\xi|})$ is small in some H^s-Sobolev norm, $s < q$. These quasi-periodic solutions are linearly stable.

This result is deduced from Theorem 5.1. It has been announced in [3]. Let us make some comments.

1. The set of tangential sites S satisfying ($\mathtt{S}1$)–($\mathtt{S}2$) can be iteratively constructed in an explicit way, see the end of Section 9. After fixing $\{\bar{\jmath}_1, \dots, \bar{\jmath}_n\}$, in the choice of $\bar{\jmath}_{n+1}$ there are only finitely many forbidden values, while all the other infinitely many values are good choices for $\bar{\jmath}_{n+1}$. In this precise sense the set S is "generic".
2. The linear stability of the quasi-periodic solutions is discussed after (9.41). In a suitable set of symplectic coordinates (ψ, η, w), $\psi \in \mathbb{T}^\nu$, near the invariant torus, the linearized equations at the quasi-periodic solutions assume the form (9.41), (9.42). Actually there is a complete KAM normal form near the invariant torus (Remark 6.5), see also [7].
3. A similar result holds for perturbed (focusing/defocussing) mKdV equations

$$u_t + u_{xxx} \pm \partial_x u^3 + \mathcal{N}_4(x, u, u_x, u_{xx}, u_{xxx}) = 0 \tag{1.10}$$

for tangential sites S which satisfy $\frac{2}{2\nu - 1} \sum_{i=1}^{\nu} \bar{\jmath}_i^2 \notin \mathbb{Z}$. If the density $f(u, u_x)$ is independent on x, the result holds for *all* the choices of the tangential sites. The KdV equation (1.1) is more difficult than (1.10) because the nonlinearity is quadratic and not cubic.

An important point is that the fourth order Birkhoff normal form of KdV and mKdV is completely integrable. The present strategy of proof — that we describe in detail below — is a rather general approach for constructing small amplitude quasi-periodic solutions of quasi-linear perturbed KdV equations. For example it could be applied to generalized KdV equations with leading nonlinearity u^p, $p \geq 4$, by using the normal form techniques of Procesi–Procesi [29,30]. A further interesting open question concerns perturbations of the finite gap solutions of KdV.

Let us describe the strategy of proof of Theorem 1.1. It involves many different arguments that are of wide applicability to other PDEs. Nevertheless we think that a unique abstract KAM theorem applicable to *all* quasi-linear PDEs could not be expected. Indeed the suitable pseudo-differential operators that are required to conjugate the highest order of the linearized operator to constant coefficients, highly depend on the PDE at hand, see the discussion after (1.11).

Weak Birkhoff normal form. Once the finite set of tangential sites S has been fixed, the first step is to perform a "weak" Birkhoff normal form (weak BNF), whose goal is to find an invariant manifold of solutions of the third order approximate KdV equation (1.1), on which the dynamics is completely integrable, see Section 3. Since the KdV nonlinearity is quadratic, two steps of weak BNF are required. The present Birkhoff map is close to the identity up to *finite dimensional* operators, see Proposition 3.1. The key advantage is that it modifies $\mathcal{N}_4$ very mildly, only up to finite dimensional operators (see for example Lemma 7.1), and thus the spectral analysis of the linearized equations (that we shall perform in Section 8) is essentially the same as if we were in the original coordinates.

The weak normal form (3.5) does not remove (or normalize) the monomials $O(z^2)$. This could be done. However, we do not perform such stronger normal form (called "partial BNF" in Kuksin–Pöschel [26] and Pöschel [28]) because the corresponding Birkhoff map is close to the identity only up to an operator of order $O(\partial_x^{-1})$, and so it would produce, in the transformed vector field $\mathcal{N}_4$, terms of order ∂_{xx} and ∂_x. A fortiori, we cannot either use the full Birkhoff normal form computed in [21] for KdV, which completely diagonalizes the fourth order terms, because such Birkhoff map is only close to the identity up to a bounded operator. For the same reason, we do not use the global nonlinear Fourier transform in [21] (Birkhoff coordinates), which is close to the Fourier transform up to smoothing operators of order $O(\partial_x^{-1})$.

The weak BNF procedure of Section 3 is sufficient to find the first nonlinear (integrable) approximation of the solutions and to extract the "frequency-to-amplitude" modulation (4.10).

In Proposition 3.1 we also remove the terms $O(v^5)$, $O(v^4 z)$ in order to have sufficiently good approximate solutions so that the Nash–Moser iteration of Section 9 will converge. This is necessary for KdV whose nonlinearity is quadratic at the origin. These further steps of Birkhoff normal form are not required if the nonlinearity is yet cubic as for mKdV, see Remark 3.5. To this aim, we choose the tangential sites S such that ($\mathtt{S}2$) holds. We also note that we assume (1.5) because we use the conservation of momentum up to the homogeneity order 5, see (2.7).

Action-angle and rescaling. At this point we introduce action-angle variables on the tangential sites (Section 4) and, after the rescaling (4.5), we look for quasi-periodic solutions of the Hamiltonian (4.9). Note that the coefficients of the normal form $\mathcal{N}$ in (4.11) depend on the angles θ, unlike the usual KAM theorems [28,23], where the whole normal form is reduced to constant coefficients. This is because the weak BNF of Section 3 did *not* normalize the quadratic terms $O(z^2)$. These terms are dealt with the "linear Birkhoff normal form" (linear BNF) in Sections 8.4, 8.5. In some sense here the "partial" Birkhoff normal form of [28] is split into the weak BNF of Section 3 and the linear BNF of Sections 8.4, 8.5.

The action-angle variables are convenient for proving the stability of the solutions.

The nonlinear functional setting and the approximate inverse. We look for a zero of the nonlinear operator (5.6), whose unknown is the embedded torus and the frequency ω is seen as an "external" parameter. The solution is obtained by a Nash–Moser iterative scheme in Sobolev scales. The key step is to construct (for ω restricted to a suitable Cantor-like set) an approximate inverse (*à la* Zehnder [35]) of the linearized operator at any approximate solution. Roughly, this means to find a linear operator which is an inverse at an exact solution. A major difficulty is that the tangential and the normal dynamics near an invariant torus are strongly coupled.

The symplectic approximate decoupling. The above difficulty is overcome by implementing the abstract procedure in Berti–Bolle [7,8] developed in order to prove existence of quasi-periodic solutions for autonomous NLW (and NLS) with a multiplicative potential. This approach reduces the search of an approximate inverse for (5.6) to the invertibility of a quasi-periodically forced PDE restricted on the normal directions. This method approximately decouples the "tangential" and the "normal" dynamics around an approximate invariant torus, introducing a suitable set of symplectic variables (ψ, η, w) near the torus, see (6.19). Note that, in the first line of (6.19), ψ is the "natural" angle variable which coordinates the torus, and, in the third line, the normal variable z is only translated by the component $z_0(\psi)$ of the torus. The second line completes this transformation to a symplectic one. The canonicity of this map is proved in [7] using the isotropy of the approximate invariant torus i_δ, see Lemma 6.3. The change of variable (6.19) brings the torus i_δ "at the origin". The advantage is that the second equation in (6.29) (which corresponds to the action variables of the torus) can be immediately solved, see (6.31). Then it remains to solve the third equation (6.32), i.e. to invert the linear operator $\mathcal{L}_\omega$. This is a quasi-periodic Hamiltonian perturbed linear Airy equation of the form

$$h \mapsto \mathcal{L}_\omega h := \Pi_S^\perp \big(\omega \cdot \partial_\varphi h + \partial_{xx}(a_1 \partial_x h) + \partial_x (a_0 h) + \partial_x \mathcal{R} h \big) , \quad \forall h \in H_S^\perp , \tag{1.11}$$

where $\mathcal{R}$ is a finite dimensional remainder. The exact form of $\mathcal{L}_\omega$ is obtained in Proposition 7.6.

Reduction of the linearized operator in the normal directions. In Section 8 we conjugate the variable coefficients operator $\mathcal{L}_\omega$ in (7.34), see (1.11), to a diagonal operator with constant coefficients which describes infinitely many harmonic oscillators

$$\dot{v}_j + \mu_j^\infty v_j = 0 , \quad \mu_j^\infty := \mathrm{i}(-m_3 j^3 + m_1 j) + r_j^\infty \in \mathrm{i}\mathbb{R} , \quad j \notin S , \tag{1.12}$$

where the constants $m_3 - 1$, $m_1 \in \mathbb{R}$ and $\sup_j |r_j^\infty|$ are small, see Theorem 8.25. The main perturbative effect to the spectrum (and the eigenfunctions) of $\mathcal{L}_\omega$ is clearly due to the term $a_1(\omega t, x)\partial_{xxx}$ (see (1.11)), and it is too strong for the usual reducibility KAM techniques to work directly. The conjugacy of $\mathcal{L}_\omega$ with (1.12) is obtained in several steps. The first task (obtained in Sections 8.1–8.6) is to conjugate $\mathcal{L}_\omega$ to another Hamiltonian operator of $H_S^\perp$ with constant coefficients

$$\mathcal{L}_6 := \Pi_S^\perp \big(\omega \cdot \partial_\varphi + m_3 \partial_{xxx} + m_1 \partial_x + R_6 \big) \Pi_S^\perp , \quad m_1, m_3 \in \mathbb{R} , \tag{1.13}$$

up to a small bounded remainder $R_6 = O(\partial_x^0)$, see (8.113). This expansion of $\mathcal{L}_\omega$ in "decreasing symbols" with constant coefficients follows [2], and it is somehow in the spirit of the works of Iooss, Plotnikov and Toland [18,17] in water waves theory, and Baldi [1] for Benjamin–Ono. It is obtained by transformations which are very different from the usual KAM changes of variables. We underline that the specific form of these transformations depend on the structure of KdV. For other quasi-linear PDEs the analogous reduction requires different transformations. For the reduction of (1.11) there are several differences with respect to [2], that we now outline.

Major differences with respect to [2] for transforming (1.11) *into* (1.13).

1. The first step is to eliminate the x-dependence from the coefficient $a_1(\omega t, x)\partial_{xxx}$ of the Hamiltonian operator $\mathcal{L}_\omega$. We cannot use the symplectic transformation $\mathcal{A}$ defined in (8.1), used in [2], because $\mathcal{L}_\omega$ acts on the normal subspace $H_S^\perp$ only, and not on the whole Sobolev space as in [2]. We cannot use the restricted map $\mathcal{A}_\perp := \Pi_S^\perp \mathcal{A} \Pi_S^\perp$, because it is *not* symplectic. In order to find a symplectic diffeomorphism of $H_S^\perp$ near $\mathcal{A}_\perp$, the first observation is to realize $\mathcal{A}$ as the flow map of the time dependent Hamiltonian transport linear PDE (8.3). Thus we conjugate $\mathcal{L}_\omega$ with the flow map of the projected Hamiltonian equation (8.5). In Lemma 8.2 we prove that it differs from $\mathcal{A}_\perp$ for finite dimensional operators. A technical, but important, fact is that the remainders produced after this conjugation of $\mathcal{L}_\omega$ remain of the finite dimensional form (7.7), see Lemma 8.3.
This step may be seen as a quantitative application of the Egorov theorem, see [31], which describes how the principal symbol of a pseudo-differential operator (here $a_1(\omega t, x)\partial_{xxx}$) transforms under the flow of a linear hyperbolic PDE (here (8.5)).
2. The operator $\mathcal{L}_\omega$ has variable coefficients also at the orders $O(\varepsilon)$ and $O(\varepsilon^2)$, see (7.34)–(7.35). This is a consequence of the fact that the weak BNF procedure of Section 3 did not touch the quadratic terms $O(z^2)$. These terms cannot be reduced to constants by the perturbative scheme in [2], which applies to terms R such that $R\gamma^{-1} \ll 1$ where γ is the diophantine constant of the frequency vector ω (the case in [2] is simpler because the diophantine constant is $\gamma = O(1)$). Here, since KdV is completely resonant, such $\gamma = o(\varepsilon^2)$, see (5.4). These terms are reduced to constant coefficients in Sections 8.4–8.5 by means of purely algebraic arguments (linear BNF), which, ultimately, stem from the complete integrability of the fourth order BNF of the KdV equation (1.6), see [21].
3. The order of the transformations of Sections 8.1–8.7 used to reduce $\mathcal{L}_\omega$ is not accidental. The first two steps in Sections 8.1, 8.2 reduce to constant coefficients the quasi-linear term $O(\partial_{xxx})$ and eliminate the term $O(\partial_{xx})$, see (8.45) (the second transformation is a time quasi-periodic reparametrization of time). Then, in Section 8.3, we apply the transformation $\mathcal{T}$ (8.64) in such a way that the space average of the coefficient $d_1(\varphi, \cdot)$ in (8.65) is constant. This is done in view of the applicability of the "descent method" in Section 8.6 where we reduce to constant coefficients the order $O(\partial_x)$ of the operator. All these transformations are composition operators induced by diffeomorphisms of the torus. Therefore they are well-defined operators of a Sobolev space into itself, but their decay norm is infinite! We perform the transformation $\mathcal{T}$ *before* the linear Birkhoff normal form steps of Sections 8.4–8.5, because $\mathcal{T}$ is a change of variable that preserves the form (7.7) of the remainders

(it is not evident after the Birkhoff normal form). The Birkhoff transformations are symplectic maps of the form $I + \varepsilon O(\partial_x^{-1})$. Thanks to this property the coefficient $d_1(\varphi, x)$ obtained in step 8.3 is *not* changed by these Birkhoff maps. The transformation in Section 8.6 is one step of "descent method" which transforms $d_1(\varphi, x)\partial_x$ into the constant coefficients differential operator $m_1 \partial_x$, up to a zero-order term $O(\partial_x^0)$. The name "descent method" has been used in Iooss–Plotnikov–Toland [16–18] to denote the iterative procedure of reduction of a linear differential (or pseudo-differential) operator with variable coefficients into one with constant coefficients, at any finite order of regularization, up to smoother remainders. The required conjugation transformations have the form $\mathrm{Id} + \sum_k a_k(\varphi, x)\partial_x^{-k}$ where the coefficients can be iteratively computed in decreasing orders. In this paper it is sufficient to stop at the order ∂_x^0, i.e. to perform only the step of Section 8.6. It is here that we use the assumption $(\mathtt{S}1)$ on the tangential sites, so that the space average of the function $q_{>2}$ is zero, see Lemma 7.5. Actually we only need that the average of the function in (7.33) is zero. If $f_5 = 0$ (see (1.5)) then $(\mathtt{S}1)$ is not required. This completes the task of conjugating $\mathcal{L}_\omega$ to $\mathcal{L}_6$ in (1.13).

Diagonalization of (1.13). Finally, in Section 8.7 we apply the abstract reducibility Theorem 4.2 in [2], based on a quadratic KAM scheme, which completely diagonalizes the linearized operator, obtaining (1.12). The required smallness condition (8.115) for R_6 holds. Indeed the biggest term in R_6 comes from the conjugation of $\varepsilon \partial_x v_\varepsilon(\theta_0(\varphi), y_\delta(\varphi))$ in (7.35). The linear BNF procedure of Section 8.4 had eliminated its main contribution $\varepsilon \partial_x v_\varepsilon(\varphi, 0)$. It remains $\varepsilon \partial_x \big(v_\varepsilon(\theta_0(\varphi), y_\delta(\varphi)) - v_\varepsilon(\varphi, 0)\big)$ which has size $O(\varepsilon^{7-2b}\gamma^{-1})$ due to the estimate (6.4) of the approximate solution. This term enters in the variable coefficients of $d_1(\varphi, x)\partial_x$ and $d_0(\varphi, x)\partial_x^0$. The first one had been reduced to the constant operator $m_1 \partial_x$ by the descent method of Section 8.6. The latter term is an operator of order $O(\partial_x^0)$ which satisfies (8.115). Thus $\mathcal{L}_6$ may be diagonalized by the iterative scheme of Theorem 4.2 in [2] which requires the smallness condition $O(\varepsilon^{7-2b}\gamma^{-2}) \ll 1$. This is the content of Section 8.7.

The Nash–Moser iteration. In Section 9 we perform the nonlinear Nash–Moser iteration which finally proves Theorem 5.1 and, therefore, Theorem 1.1. The optimal smallness condition required for the convergence of the scheme is $\varepsilon \|\mathcal{F}(\varphi, 0, 0)\|_{s_0+\mu}\gamma^{-2} \ll 1$, see (9.5). It is verified because $\|X_P(\varphi, 0, 0)\|_s \leq_s \varepsilon^{6-2b}$ (see (5.15)), which, in turn, is a consequence of having eliminated the terms $O(v^5)$, $O(v^4 z)$ from the original Hamiltonian (3.1), see (3.5). This requires the condition $(\mathtt{S}2)$.

2. Preliminaries

2.1. Hamiltonian formalism of KdV

The Hamiltonian vector field X_H generated by a Hamiltonian $H : H_0^1(\mathbb{T}_x) \to \mathbb{R}$ is $X_H(u) := \partial_x \nabla H(u)$, because

$$dH(u)[h] = (\nabla H(u), h)_{L^2(\mathbb{T}_x)} = \Omega(X_H(u), h), \quad \forall u, h \in H_0^1(\mathbb{T}_x),$$

where Ω is the non-degenerate symplectic form

$$\Omega(u, v) := \int_{\mathbb{T}} (\partial_x^{-1} u)\, v \, dx, \quad \forall u, v \in H_0^1(\mathbb{T}_x), \tag{2.1}$$

and $\partial_x^{-1} u$ is the periodic primitive of u with zero average. Note that

$$\partial_x \partial_x^{-1} = \partial_x^{-1}\partial_x = \pi_0, \quad \pi_0(u) := u - \frac{1}{2\pi}\int_{\mathbb{T}} u(x)\, dx. \tag{2.2}$$

A map is symplectic if it preserves the 2-form Ω.

We also remind that the Poisson bracket between two functions $F, G : H_0^1(\mathbb{T}_x) \to \mathbb{R}$ is

$$\{F(u), G(u)\} := \Omega(X_F, X_G) = \int_{\mathbb{T}} \nabla F(u)\partial_x \nabla G(u) dx. \tag{2.3}$$

The linearized KdV equation at u is

$$h_t = \partial_x (\partial_u \nabla H)(u)[h] = X_K(h),$$

where X_K is the KdV Hamiltonian vector field with quadratic Hamiltonian $K = \frac{1}{2}((\partial_u \nabla H)(u)[h], h)_{L^2(\mathbb{T}_x)} = \frac{1}{2}(\partial_{uu} H)(u)[h, h]$. By the Schwartz theorem, the Hessian operator $A := (\partial_u \nabla H)(u)$ is symmetric, namely $A^T = A$, with respect to the L^2-scalar product.

Dynamical systems formulation. It is convenient to regard the KdV equation also in the Fourier representation

$$u(x) = \sum_{j\in\mathbb{Z}\setminus\{0\}} u_j e^{\mathrm{i}jx}, \qquad u(x) \longleftrightarrow u := (u_j)_{j\in\mathbb{Z}\setminus\{0\}}, \quad u_{-j} = \bar{u}_j, \tag{2.4}$$

where the Fourier indices $j \in \mathbb{Z}\setminus\{0\}$ by the definition (1.4) of the phase space and $u_{-j} = \bar{u}_j$ because $u(x)$ is real-valued. The symplectic structure writes

$$\Omega = \frac{1}{2}\sum_{j\neq 0} \frac{1}{\mathrm{i}j} du_j \wedge du_{-j} = \sum_{j\geq 1} \frac{1}{\mathrm{i}j} du_j \wedge du_{-j}, \qquad \Omega(u, v) = \sum_{j\neq 0} \frac{1}{\mathrm{i}j} u_j v_{-j} = \sum_{j\neq 0} \frac{1}{\mathrm{i}j} u_j \bar{v}_j, \tag{2.5}$$

the Hamiltonian vector field X_H and the Poisson bracket $\{F, G\}$ are

$$[X_H(u)]_j = \mathrm{i}j(\partial_{u_{-j}} H)(u), \ \forall j \neq 0, \quad \{F(u), G(u)\} = -\sum_{j\neq 0} \mathrm{i}j(\partial_{u_{-j}} F)(u)(\partial_{u_j} G)(u). \tag{2.6}$$

Conservation of momentum. A Hamiltonian

$$H(u) = \sum_{j_1,\dots,j_n\in\mathbb{Z}\setminus\{0\}} H_{j_1,\dots,j_n} u_{j_1} \dots u_{j_n}, \quad u(x) = \sum_{j\in\mathbb{Z}\setminus\{0\}} u_j e^{\mathrm{i}jx}, \tag{2.7}$$

homogeneous of degree n, *preserves the momentum* if the coefficients $H_{j_1,\dots,j_n}$ are zero for $j_1 + \dots + j_n \neq 0$, so that the sum in (2.7) is restricted to integers such that $j_1 + \dots + j_n = 0$. Equivalently, H preserves the momentum if $\{H, M\} = 0$, where M is the momentum $M(u) := \int_{\mathbb{T}} u^2 dx = \sum_{j\in\mathbb{Z}\setminus\{0\}} u_j u_{-j}$. The homogeneous components of degree ≤ 5 of the KdV Hamiltonian H in (1.3) preserve the momentum because, by (1.5), the homogeneous component f_5 of degree 5 does not depend on the space variable x.

Tangential and normal variables. Let $\bar{\jmath}_1, \dots, \bar{\jmath}_\nu \geq 1$ be ν distinct integers, and $S^+ := \{\bar{\jmath}_1, \dots, \bar{\jmath}_\nu\}$. Let S be the symmetric set in (1.8), and $S^c := \{j \in \mathbb{Z}\setminus\{0\} : j \notin S\}$ its complementary set in $\mathbb{Z}\setminus\{0\}$. We decompose the phase space as

$$H_0^1(\mathbb{T}_x) := H_S \oplus H_S^\perp, \quad H_S := \mathrm{span}\{e^{\mathrm{i}jx} : j \in S\}, \quad H_S^\perp := \Big\{u = \sum_{j\in S^c} u_j e^{\mathrm{i}jx} \in H_0^1(\mathbb{T}_x)\Big\}, \tag{2.8}$$

and we denote by Π_S, $\Pi_S^\perp$ the corresponding orthogonal projectors. Accordingly we decompose

$$u = v + z, \qquad v = \Pi_S u := \sum_{j\in S} u_j\, e^{\mathrm{i}jx}, \quad z = \Pi_S^\perp u := \sum_{j\in S^c} u_j\, e^{\mathrm{i}jx}, \tag{2.9}$$

where v is called the *tangential* variable and z the *normal* one. We shall sometimes identify $v \equiv (v_j)_{j\in S}$ and $z \equiv (z_j)_{j\in S^c}$. The subspaces H_S and $H_S^\perp$ are *symplectic*. The dynamics of these two components is quite different. On H_S we shall introduce the action-angle variables, see (4.1). The linear frequencies of oscillations on the tangential sites are

$$\bar{\omega} := (\bar{\jmath}_1^3, \dots, \bar{\jmath}_\nu^3) \in \mathbb{N}^\nu. \tag{2.10}$$

2.2. *Functional setting*

Norms. Along the paper we shall use the notation

$$\|u\|_s := \|u\|_{H^s(\mathbb{T}^{\nu+1})} := \|u\|_{H^s_{\varphi,x}} \tag{2.11}$$

to denote the Sobolev norm of functions $u = u(\varphi, x)$ in the Sobolev space $H^s(\mathbb{T}^{\nu+1})$. We shall denote by $\| \ \|_{H^s_x}$ the Sobolev norm in the phase space of functions $u := u(x) \in H^s(\mathbb{T})$. Moreover $\| \ \|_{H^s_\varphi}$ will denote the Sobolev norm of scalar functions, like the Fourier components $u_j(\varphi)$.

We fix $s_0 := (\nu+2)/2$ so that $H^{s_0}(\mathbb{T}^{\nu+1}) \hookrightarrow L^\infty(\mathbb{T}^{\nu+1})$ and the spaces $H^s(\mathbb{T}^{\nu+1})$, $s > s_0$, are an algebra. At the end of this section we report interpolation properties of the Sobolev norm that will be currently used along the paper. We shall also denote

$$H^s_{S^\perp}(\mathbb{T}^{\nu+1}) := \big\{ u \in H^s(\mathbb{T}^{\nu+1}) \, : \, u(\varphi, \cdot) \in H_S^\perp \; \forall \varphi \in \mathbb{T}^\nu \big\}\,, \tag{2.12}$$
$$H^s_S(\mathbb{T}^{\nu+1}) := \big\{ u \in H^s(\mathbb{T}^{\nu+1}) \, : \, u(\varphi, \cdot) \in H_S \; \forall \varphi \in \mathbb{T}^\nu \big\}\,. \tag{2.13}$$

For a function $u : \Omega_o \to E$, $\omega \mapsto u(\omega)$, where $(E, \| \, \|_E)$ is a Banach space and Ω_o is a subset of $\mathbb{R}^\nu$, we define the sup-norm and the Lipschitz semi-norm

$$\|u\|^{\sup}_E := \|u\|^{\sup}_{E,\Omega_o} := \sup_{\omega \in \Omega_o} \|u(\omega)\|_E\,, \quad \|u\|^{\mathrm{lip}}_E := \|u\|^{\mathrm{lip}}_{E,\Omega_o} := \sup_{\omega_1 \neq \omega_2} \frac{\|u(\omega_1) - u(\omega_2)\|_E}{|\omega_1 - \omega_2|}\,, \tag{2.14}$$

and, for $\gamma > 0$, the Lipschitz norm

$$\|u\|^{\mathrm{Lip}(\gamma)}_E := \|u\|^{\mathrm{Lip}(\gamma)}_{E,\Omega_o} := \|u\|^{\sup}_E + \gamma \|u\|^{\mathrm{lip}}_E\,. \tag{2.15}$$

If $E = H^s$ we simply denote $\|u\|^{\mathrm{Lip}(\gamma)}_{H^s} := \|u\|^{\mathrm{Lip}(\gamma)}_s$. We shall use the notation

$$a \leq_s b \quad \Longleftrightarrow \quad a \leq C(s) b \quad \text{for some constant } C(s) > 0\,.$$

Matrices with off-diagonal decay. A linear operator can be identified, as usual, with its matrix representation. We recall the definition of the s-decay norm (introduced in [6]) of an infinite dimensional matrix. This norm is used in [2] for the KAM reducibility scheme of the linearized operators.

Definition 2.1. The s-decay norm of an infinite dimensional matrix $A := (A^{i_2}_{i_1})_{i_1, i_2 \in \mathbb{Z}^b}$, $b \geq 1$, is

$$|A|^2_s := \sum_{i \in \mathbb{Z}^b} \langle i \rangle^{2s} \Big(\sup_{i_1 - i_2 = i} |A^{i_2}_{i_1}| \Big)^2\,. \tag{2.16}$$

For parameter dependent matrices $A := A(\omega)$, $\omega \in \Omega_o \subseteq \mathbb{R}^\nu$, the definitions (2.14) and (2.15) become

$$|A|^{\sup}_s := \sup_{\omega \in \Omega_o} |A(\omega)|_s\,, \quad |A|^{\mathrm{lip}}_s := \sup_{\omega_1 \neq \omega_2} \frac{|A(\omega_1) - A(\omega_2)|_s}{|\omega_1 - \omega_2|}\,, \quad |A|^{\mathrm{Lip}(\gamma)}_s := |A|^{\sup}_s + \gamma |A|^{\mathrm{lip}}_s\,. \tag{2.17}$$

Such a norm is modeled on the behavior of matrices representing the multiplication operator by a function. Actually, given a function $p \in H^s(\mathbb{T}^b)$, the multiplication operator $h \mapsto ph$ is represented by the Töplitz matrix $T^{i'}_i = p_{i-i'}$ and $|T|_s = \|p\|_s$. If $p = p(\omega)$ is a Lipschitz family of functions, then

$$|T|^{\mathrm{Lip}(\gamma)}_s = \|p\|^{\mathrm{Lip}(\gamma)}_s\,. \tag{2.18}$$

The s-norm satisfies classical algebra and interpolation inequalities, see [2].

Lemma 2.1. *Let $A = A(\omega)$ and $B = B(\omega)$ be matrices depending in a Lipschitz way on the parameter $\omega \in \Omega_o \subset \mathbb{R}^\nu$. Then for all $s \geq s_0 > b/2$ there are $C(s) \geq C(s_0) \geq 1$ such that*

$$|AB|^{\mathrm{Lip}(\gamma)}_s \leq C(s) |A|^{\mathrm{Lip}(\gamma)}_s |B|^{\mathrm{Lip}(\gamma)}_s\,, \tag{2.19}$$
$$|AB|^{\mathrm{Lip}(\gamma)}_s \leq C(s) |A|^{\mathrm{Lip}(\gamma)}_s |B|^{\mathrm{Lip}(\gamma)}_{s_0} + C(s_0) |A|^{\mathrm{Lip}(\gamma)}_{s_0} |B|^{\mathrm{Lip}(\gamma)}_s\,. \tag{2.20}$$

The s-decay norm controls the Sobolev norm, namely

$$\|Ah\|^{\mathrm{Lip}(\gamma)}_s \leq C(s) \big(|A|^{\mathrm{Lip}(\gamma)}_{s_0} \|h\|^{\mathrm{Lip}(\gamma)}_s + |A|^{\mathrm{Lip}(\gamma)}_s \|h\|^{\mathrm{Lip}(\gamma)}_{s_0} \big)\,. \tag{2.21}$$

Let now $b := \nu + 1$. An important sub-algebra is formed by the *Töplitz in time matrices* defined by

$$A^{(l_2, j_2)}_{(l_1, j_1)} := A^{j_2}_{j_1}(l_1 - l_2)\,, \tag{2.22}$$

whose decay norm (2.16) is

$$|A|_s^2 = \sum_{j \in \mathbb{Z}, l \in \mathbb{Z}^\nu} \big(\sup_{j_1 - j_2 = j} |A_{j_1}^{j_2}(l)| \big)^2 \langle l, j \rangle^{2s} . \tag{2.23}$$

These matrices are identified with the φ-dependent family of operators

$$A(\varphi) := \big(A_{j_1}^{j_2}(\varphi) \big)_{j_1, j_2 \in \mathbb{Z}}, \quad A_{j_1}^{j_2}(\varphi) := \sum_{l \in \mathbb{Z}^\nu} A_{j_1}^{j_2}(l) e^{\mathrm{i} l \cdot \varphi} \tag{2.24}$$

which act on functions of the x-variable as

$$A(\varphi) : h(x) = \sum_{j \in \mathbb{Z}} h_j e^{\mathrm{i} j x} \mapsto A(\varphi) h(x) = \sum_{j_1, j_2 \in \mathbb{Z}} A_{j_1}^{j_2}(\varphi) h_{j_2} e^{\mathrm{i} j_1 x} . \tag{2.25}$$

We still denote by $|A(\varphi)|_s$ the s-decay norm of the matrix in (2.24). As in [2], all the transformations that we shall construct in this paper are of this type (with $j, j_1, j_2 \neq 0$ because they act on the phase space $H_0^1(\mathbb{T}_x)$). This observation allows to interpret the conjugacy procedure from a dynamical point of view, see [2]-Section 2.2. Let us fix some terminology.

Definition 2.2. We say that:

the operator $(Ah)(\varphi, x) := A(\varphi) h(\varphi, x)$ is SYMPLECTIC if each $A(\varphi)$, $\varphi \in \mathbb{T}^\nu$, is a symplectic map of the phase space (or of a symplectic subspace like $H_S^\perp$);

the operator $\omega \cdot \partial_\varphi - \partial_x G(\varphi)$ is HAMILTONIAN if each $G(\varphi)$, $\varphi \in \mathbb{T}^\nu$, is symmetric;

an operator is REAL if it maps real-valued functions into real-valued functions.

As well known, a Hamiltonian operator $\omega \cdot \partial_\varphi - \partial_x G(\varphi)$ is transformed, under a symplectic map $\mathcal{A}$, into another Hamiltonian operator $\omega \cdot \partial_\varphi - \partial_x E(\varphi)$, see e.g. [2]-Section 2.3.

We conclude this preliminary section recalling the following well known lemmata, see Appendix of [2].

Lemma 2.2 *(Composition). Assume $f \in C^s(\mathbb{T}^d \times B_1)$, $B_1 := \{y \in \mathbb{R}^m : |y| \le 1\}$. Then $\forall u \in H^s(\mathbb{T}^d, \mathbb{R}^m)$ such that $\|u\|_{L^\infty} < 1$, the composition operator $\tilde f(u)(x) := f(x, u(x))$ satisfies $\|\tilde f(u)\|_s \le C \|f\|_{C^s} (\|u\|_s + 1)$ where the constant C depends on s, d. If $f \in C^{s+2}$ and $\|u + h\|_{L^\infty} < 1$, then*

$$\Big\| \tilde f(u+h) - \sum_{i=0}^{k} \frac{\tilde f^{(i)}(u)}{i!} [h^i] \Big\|_s \le C \|f\|_{C^{s+2}} \|h\|_{L^\infty}^k (\|h\|_s + \|h\|_{L^\infty} \|u\|_s), \quad k = 0, 1 .$$

The previous statement also holds replacing $\| \ \|_s$ with the norms $| \ |_{s,\infty}$.

Lemma 2.3 *(Tame product). For $s \ge s_0 > d/2$,*

$$\|uv\|_s \le C(s_0) \|u\|_s \|v\|_{s_0} + C(s) \|u\|_{s_0} \|v\|_s , \quad \forall u, v \in H^s(\mathbb{T}^d) .$$

For $s \ge 0$, $s \in \mathbb{N}$,

$$\|uv\|_s \le \tfrac{3}{2} \|u\|_{L^\infty} \|v\|_s + C(s) \|u\|_{W^{s,\infty}} \|v\|_0 , \quad \forall u \in W^{s,\infty}(\mathbb{T}^d) , \ v \in H^s(\mathbb{T}^d) .$$

The above inequalities also hold for the norms $\| \ \|_s^{\mathrm{Lip}(\gamma)}$.

Lemma 2.4 *(Change of variable). Let $p \in W^{s,\infty}(\mathbb{T}^d, \mathbb{R}^d)$, $s \ge 1$, with $\|p\|_{W^{1,\infty}} \le 1/2$. Then the function $f(x) = x + p(x)$ is invertible, with inverse $f^{-1}(y) = y + q(y)$ where $q \in W^{s,\infty}(\mathbb{T}^d, \mathbb{R}^d)$, and $\|q\|_{W^{s,\infty}} \le C \|p\|_{W^{s,\infty}}$. If, moreover, $p = p_\omega$ depends in a Lipschitz way on a parameter $\omega \in \Omega \subset \mathbb{R}^\nu$, and $\|D_x p_\omega\|_{L^\infty} \le 1/2$, $\forall \omega$, then $\|q\|_{W^{s,\infty}}^{\mathrm{Lip}(\gamma)} \le C \|p\|_{W^{s+1,\infty}}^{\mathrm{Lip}(\gamma)}$. The constant $C := C(d, s)$ is independent of γ.*

If $u \in H^s(\mathbb{T}^d, \mathbb{C})$, then $(u \circ f)(x) := u(x + p(x))$ satisfies

$$\|u \circ f\|_s \le C(\|u\|_s + \|p\|_{W^{s,\infty}} \|u\|_1), \quad \|u \circ f - u\|_s \le C(\|p\|_{L^\infty} \|u\|_{s+1} + \|p\|_{W^{s,\infty}} \|u\|_2),$$
$$\|u \circ f\|_s^{\mathrm{Lip}(\gamma)} \le C \big(\|u\|_{s+1}^{\mathrm{Lip}(\gamma)} + \|p\|_{W^{s,\infty}}^{\mathrm{Lip}(\gamma)} \|u\|_2^{\mathrm{Lip}(\gamma)} \big).$$

The function $u \circ f^{-1}$ satisfies the same bounds.

3. Weak Birkhoff normal form

The Hamiltonian of the perturbed KdV equation (1.1) is $H = H_2 + H_3 + H_{\geq 5}$ (see (1.3)) where

$$H_2(u) := \frac{1}{2}\int_{\mathbb{T}} u_x^2\, dx\,, \quad H_3(u) := \int_{\mathbb{T}} u^3 dx\,, \quad H_{\geq 5}(u) := \int_{\mathbb{T}} f(x,u,u_x)dx\,, \tag{3.1}$$

and f satisfies (1.5). According to the splitting (2.9) $u = v + z$, $v \in H_S$, $z \in H_S^{\perp}$, we have

$$H_2(u) = \int_{\mathbb{T}} \frac{v_x^2}{2}\, dx + \int_{\mathbb{T}} \frac{z_x^2}{2}\, dx, \quad H_3(u) = \int_{\mathbb{T}} v^3 dx + 3\int_{\mathbb{T}} v^2 z dx + 3\int_{\mathbb{T}} v z^2 dx + \int_{\mathbb{T}} z^3 dx\,. \tag{3.2}$$

For a finite-dimensional space

$$E := E_C := \mathrm{span}\{e^{\mathrm{i}jx} : 0 < |j| \leq C\}, \quad C > 0, \tag{3.3}$$

let Π_E denote the corresponding L^2-projector on E.

The notation $R(v^{k-q}z^q)$ indicates a homogeneous polynomial of degree k in (v,z) of the form

$$R(v^{k-q}z^q) = M[\underbrace{v,\ldots,v}_{(k-q)\text{ times}}, \underbrace{z,\ldots,z}_{q\text{ times}}], \qquad M = k\text{-linear}\,.$$

Proposition 3.1 *(Weak Birkhoff normal form). Assume Hypothesis* (S2). *Then there exists an analytic invertible symplectic transformation of the phase space* $\Phi_B : H_0^1(\mathbb{T}_x) \to H_0^1(\mathbb{T}_x)$ *of the form*

$$\Phi_B(u) = u + \Psi(u), \quad \Psi(u) = \Pi_E \Psi(\Pi_E u), \tag{3.4}$$

where E is a finite-dimensional space as in (3.3), such that the transformed Hamiltonian is

$$\mathcal{H} := H \circ \Phi_B = H_2 + \mathcal{H}_3 + \mathcal{H}_4 + \mathcal{H}_5 + \mathcal{H}_{\geq 6}\,, \tag{3.5}$$

where H_2 is defined in (3.1),

$$\mathcal{H}_3 := \int_{\mathbb{T}} z^3\, dx + 3\int_{\mathbb{T}} v z^2\, dx\,, \quad \mathcal{H}_4 := -\frac{3}{2}\sum_{j\in S}\frac{|u_j|^4}{j^2} + \mathcal{H}_{4,2} + \mathcal{H}_{4,3}\,, \quad \mathcal{H}_5 := \sum_{q=2}^{5} R(v^{5-q}z^q)\,, \tag{3.6}$$

$$\mathcal{H}_{4,2} := 6\int_{\mathbb{T}} v z \Pi_S\big((\partial_x^{-1}v)(\partial_x^{-1}z)\big)\, dx + 3\int_{\mathbb{T}} z^2 \pi_0(\partial_x^{-1}v)^2\, dx\,, \quad \mathcal{H}_{4,3} := R(vz^3)\,, \tag{3.7}$$

and $\mathcal{H}_{\geq 6}$ collects all the terms of order at least six in (v,z).

The rest of this section is devoted to the proof of Proposition 3.1.

First, we remove the cubic terms $\int_{\mathbb{T}} v^3 + 3\int_{\mathbb{T}} v^2 z$ from the Hamiltonian H_3 defined in (3.2). In the Fourier coordinates (2.4), we have

$$H_2 = \frac{1}{2}\sum_{j\neq 0} j^2 |u_j|^2, \quad H_3 = \sum_{j_1+j_2+j_3=0} u_{j_1}u_{j_2}u_{j_3}\,. \tag{3.8}$$

We look for a symplectic transformation $\Phi^{(3)}$ of the phase space which eliminates the monomials $u_{j_1}u_{j_2}u_{j_3}$ of H_3 with at most *one* index outside S. Note that, by the relation $j_1 + j_2 + j_3 = 0$, they are *finitely* many. We look for $\Phi^{(3)} := (\Phi^t_{F^{(3)}})_{|t=1}$ as the time-1 flow map generated by the Hamiltonian vector field $X_{F^{(3)}}$, with an auxiliary Hamiltonian of the form

$$F^{(3)}(u) := \sum_{j_1+j_2+j_3=0} F^{(3)}_{j_1 j_2 j_3} u_{j_1}u_{j_2}u_{j_3}\,.$$

The transformed Hamiltonian is

$$H^{(3)} := H \circ \Phi^{(3)} = H_2 + H_3^{(3)} + H_4^{(3)} + H_{\geq 5}^{(3)},$$
$$H_3^{(3)} = H_3 + \{H_2, F^{(3)}\}, \quad H_4^{(3)} = \frac{1}{2}\{\{H_2, F^{(3)}\}, F^{(3)}\} + \{H_3, F^{(3)}\}, \tag{3.9}$$

where $H_{\geq 5}^{(3)}$ collects all the terms of order at least five in (u, u_x). By (3.8) and (2.6) we calculate

$$H_3^{(3)} = \sum_{j_1+j_2+j_3=0} \{1 - \mathrm{i}(j_1^3 + j_2^3 + j_3^3) F_{j_1 j_2 j_3}^{(3)}\} u_{j_1} u_{j_2} u_{j_3}.$$

Hence, in order to eliminate the monomials with at most one index outside S, we choose

$$F_{j_1 j_2 j_3}^{(3)} := \begin{cases} \frac{1}{\mathrm{i}(j_1^3+j_2^3+j_3^3)} & \text{if } (j_1, j_2, j_3) \in \mathcal{A}, \\ 0 & \text{otherwise}, \end{cases} \tag{3.10}$$

where $\mathcal{A} := \{(j_1, j_2, j_3) \in (\mathbb{Z} \setminus \{0\})^3 : j_1 + j_2 + j_3 = 0,\ j_1^3 + j_2^3 + j_3^3 \neq 0$, and at least 2 among j_1, j_2, j_3 belong to $S\}$. Note that

$$\mathcal{A} = \{(j_1, j_2, j_3) \in (\mathbb{Z} \setminus \{0\})^3 : j_1 + j_2 + j_3 = 0, \text{ and at least 2 among } j_1, j_2, j_3 \text{ belong to } S\} \tag{3.11}$$

because of the elementary relation

$$j_1 + j_2 + j_3 = 0 \quad \Rightarrow \quad j_1^3 + j_2^3 + j_3^3 = 3 j_1 j_2 j_3 \neq 0 \tag{3.12}$$

being $j_1, j_2, j_3 \in \mathbb{Z} \setminus \{0\}$. Also note that $\mathcal{A}$ is a finite set, actually $\mathcal{A} \subseteq [-2C_S, 2C_S]^3$ where the tangential sites $S \subseteq [-C_S, C_S]$. As a consequence, the Hamiltonian vector field $X_{F^{(3)}}$ has finite rank and vanishes outside the finite dimensional subspace $E := E_{2C_S}$ (see (3.3)), namely

$$X_{F^{(3)}}(u) = \Pi_E X_{F^{(3)}}(\Pi_E u).$$

Hence its flow $\Phi^{(3)} : H_0^1(\mathbb{T}_x) \to H_0^1(\mathbb{T}_x)$ has the form (3.4) and it is analytic.

By construction, all the monomials of H_3 with at least two indices outside S are not modified by the transformation $\Phi^{(3)}$. Hence (see (3.2)) we have

$$H_3^{(3)} = \int_{\mathbb{T}} z^3\, dx + 3 \int_{\mathbb{T}} v z^2\, dx. \tag{3.13}$$

We now compute the fourth order term $H_4^{(3)} = \sum_{i=0}^4 H_{4,i}^{(3)}$ in (3.9), where $H_{4,i}^{(3)}$ is of type $R(v^{4-i} z^i)$.

Lemma 3.2. *One has (recall the definition (2.2) of π_0)*

$$H_{4,0}^{(3)} := \frac{3}{2} \int_{\mathbb{T}} v^2 \pi_0[(\partial_x^{-1} v)^2] dx, \quad H_{4,2}^{(3)} := 6 \int_{\mathbb{T}} v z \Pi_S\big((\partial_x^{-1} v)(\partial_x^{-1} z)\big)\, dx + 3 \int_{\mathbb{T}} z^2 \pi_0[(\partial_x^{-1} v)^2] dx. \tag{3.14}$$

Proof. We write $H_3 = H_{3,\leq 1} + H_3^{(3)}$ where $H_{3,\leq 1}(u) := \int_{\mathbb{T}} v^3 dx + 3 \int_{\mathbb{T}} v^2 z\, dx$. Then, by (3.9), we get

$$H_4^{(3)} = \frac{1}{2}\{H_{3,\leq 1}, F^{(3)}\} + \{H_3^{(3)}, F^{(3)}\}. \tag{3.15}$$

By (3.10), (3.12), the auxiliary Hamiltonian may be written as

$$F^{(3)}(u) = -\frac{1}{3} \sum_{(j_1, j_2, j_3) \in \mathcal{A}} \frac{u_{j_1} u_{j_2} u_{j_3}}{(\mathrm{i} j_1)(\mathrm{i} j_2)(\mathrm{i} j_3)} = -\frac{1}{3} \int_{\mathbb{T}} (\partial_x^{-1} v)^3 dx - \int_{\mathbb{T}} (\partial_x^{-1} v)^2 (\partial_x^{-1} z) dx.$$

Hence, using that the projectors Π_S, $\Pi_S^{\perp}$ are self-adjoint and ∂_x^{-1} is skew-selfadjoint,

$$\nabla F^{(3)}(u) = \partial_x^{-1}\{(\partial_x^{-1} v)^2 + 2\Pi_S[(\partial_x^{-1} v)(\partial_x^{-1} z)]\} \tag{3.16}$$

(we have used that $\partial_x^{-1}\pi_0 = \partial_x^{-1}$ be the definition of ∂_x^{-1}). Recalling the Poisson bracket definition (2.3), using that $\nabla H_{3,\le 1}(u) = 3v^2 + 6\Pi_S(vz)$ and (3.16), we get

$$\begin{aligned}\{H_{3,\le 1}, F^{(3)}\} &= \int_{\mathbb{T}} \{3v^2 + 6\Pi_S(vz)\}\pi_0\{(\partial_x^{-1}v)^2 + 2\Pi_S[(\partial_x^{-1}v)(\partial_x^{-1}z)]\}\,dx \\ &= 3\int_{\mathbb{T}} v^2\pi_0(\partial_x^{-1}v)^2\,dx + 12\int_{\mathbb{T}} \Pi_S(vz)\Pi_S[(\partial_x^{-1}v)(\partial_x^{-1}z)]\,dx + R(v^3 z)\,. \end{aligned} \tag{3.17}$$

Similarly, since $\nabla H_3^{(3)}(u) = 3z^2 + 6\Pi_S^{\perp}(vz)$,

$$\{H_3^{(3)}, F^{(3)}\} = 3\int_{\mathbb{T}} z^2\pi_0(\partial_x^{-1}v)^2\,dx + R(v^3 z) + R(vz^3)\,. \tag{3.18}$$

The lemma follows by (3.15), (3.17), (3.18). □

We now construct a symplectic map $\Phi^{(4)}$ such that the Hamiltonian system obtained transforming $H_2 + H_3^{(3)} + H_4^{(3)}$ possesses the invariant subspace H_S (see (2.8)) and its dynamics on H_S is integrable and non-isochronous. Hence we have to eliminate the term $H_{4,1}^{(3)}$ (which is linear in z), and to normalize $H_{4,0}^{(3)}$ (which is independent of z). We need the following elementary lemma (Lemma 13.4 in [21]).

Lemma 3.3. *Let $j_1, j_2, j_3, j_4 \in \mathbb{Z}$ such that $j_1 + j_2 + j_3 + j_4 = 0$. Then*

$$j_1^3 + j_2^3 + j_3^3 + j_4^3 = -3(j_1 + j_2)(j_1 + j_3)(j_2 + j_3).$$

Lemma 3.4. *There exists a symplectic transformation $\Phi^{(4)}$ of the form (3.4) such that*

$$H^{(4)} := H^{(3)} \circ \Phi^{(4)} = H_2 + H_3^{(3)} + H_4^{(4)} + H_{\ge 5}^{(4)}\,, \qquad H_4^{(4)} := -\frac{3}{2}\sum_{j\in S}\frac{|u_j|^4}{j^2} + H_{4,2}^{(3)} + H_{4,3}^{(3)}\,, \tag{3.19}$$

where $H_3^{(3)}$ is defined in (3.13), $H_{4,2}^{(3)}$ in (3.14), $H_{4,3}^{(3)} = R(vz^3)$ and $H_{\ge 5}^{(4)}$ collects all the terms of degree at least five in (u, u_x).

Proof. We look for a map $\Phi^{(4)} := (\Phi^t_{F^{(4)}})_{|t=1}$ which is the time 1-flow map of an auxiliary Hamiltonian

$$F^{(4)}(u) := \sum_{\substack{j_1+j_2+j_3+j_4=0 \\ \text{at least 3 indices are in } S}} F^{(4)}_{j_1 j_2 j_3 j_4} u_{j_1} u_{j_2} u_{j_3} u_{j_4}$$

with the same form of the Hamiltonian $H_{4,0}^{(3)} + H_{4,1}^{(3)}$. The transformed Hamiltonian is

$$H^{(4)} := H^{(3)} \circ \Phi^{(4)} = H_2 + H_3^{(3)} + H_4^{(4)} + H_{\ge 5}^{(4)}, \quad H_4^{(4)} = \{H_2, F^{(4)}\} + H_4^{(3)}, \tag{3.20}$$

where $H_{\ge 5}^{(4)}$ collects all the terms of order at least five. We write $H_4^{(4)} = \sum_{i=0}^4 H_{4,i}^{(4)}$ where each $H_{4,i}^{(4)}$ if of type $R(v^{4-i}z^i)$. We choose the coefficients

$$F^{(4)}_{j_1 j_2 j_3 j_4} := \begin{cases} \dfrac{H^{(3)}_{j_1 j_2 j_3 j_4}}{\mathrm{i}(j_1^3 + j_2^3 + j_3^3 + j_4^3)} & \text{if } (j_1, j_2, j_3, j_4) \in \mathcal{A}_4\,, \\ 0 & \text{otherwise,} \end{cases} \tag{3.21}$$

where

$$\begin{aligned}\mathcal{A}_4 := \big\{&(j_1, j_2, j_3, j_4) \in (\mathbb{Z}\setminus\{0\})^4 : j_1 + j_2 + j_3 + j_4 = 0,\ j_1^3 + j_2^3 + j_3^3 + j_4^3 \ne 0, \\ &\text{and at most one among } j_1, j_2, j_3, j_4 \text{ outside } S\big\}\,.\end{aligned}$$

By this definition $H^{(4)}_{4,1} = 0$ because there exist no integers $j_1, j_2, j_3 \in S$, $j_4 \in S^c$ satisfying $j_1 + j_2 + j_3 + j_4 = 0$, $j_1^3 + j_2^3 + j_3^3 + j_4^3 = 0$, by Lemma 3.3 and the fact that S is symmetric. By construction, the terms $H^{(4)}_{4,i} = H^{(3)}_{4,i}$, $i = 2, 3, 4$, are not changed by $\Phi^{(4)}$. Finally, by (3.14)

$$H^{(4)}_{4,0} = \frac{3}{2} \sum_{\substack{j_1, j_2, j_3, j_4 \in S \\ j_1+j_2+j_3+j_4=0 \\ j_1^3+j_2^3+j_3^3+j_4^3=0 \\ j_1+j_2\,,\, j_3+j_4 \neq 0}} \frac{1}{(\mathrm{i} j_3)(\mathrm{i} j_4)} u_{j_1} u_{j_2} u_{j_3} u_{j_4} \,. \tag{3.22}$$

If $j_1 + j_2 + j_3 + j_4 = 0$ and $j_1^3 + j_2^3 + j_3^3 + j_4^3 = 0$, then $(j_1 + j_2)(j_1 + j_3)(j_2 + j_3) = 0$ by Lemma 3.3. We develop the sum in (3.22) with respect to the first index j_1. Since $j_1 + j_2 \neq 0$ the possible cases are:

$$(i)\ \big\{j_2 \neq -j_1,\ j_3 = -j_1,\ j_4 = -j_2\big\} \qquad \text{or} \qquad (ii)\ \big\{j_2 \neq -j_1,\ j_3 \neq -j_1,\ j_3 = -j_2,\ j_4 = -j_1\big\}.$$

Hence, using $u_{-j} = \bar{u}_j$ (recall (2.4)), and since S is symmetric, we have

$$\sum_{(i)} \frac{1}{j_3 j_4} u_{j_1} u_{j_2} u_{j_3} u_{j_4} = \sum_{j_1, j_2 \in S, j_2 \neq -j_1} \frac{|u_{j_1}|^2 |u_{j_2}|^2}{j_1 j_2} = \sum_{j, j' \in S} \frac{|u_j|^2 |u_{j'}|^2}{j j'} + \sum_{j \in S} \frac{|u_j|^4}{j^2} = \sum_{j \in S} \frac{|u_j|^4}{j^2}\,, \tag{3.23}$$

and in the second case (ii)

$$\sum_{(ii)} \frac{1}{j_3 j_4} u_{j_1} u_{j_2} u_{j_3} u_{j_4} = \sum_{j_1, j_2, j_2 \neq \pm j_1} \frac{1}{j_1 j_2} u_{j_1} u_{j_2} u_{-j_2} u_{-j_1} = \sum_{j \in S} \frac{1}{j} |u_j|^2 \Big(\sum_{j_2 \neq \pm j} \frac{1}{j_2} |u_{j_2}|^2 \Big) = 0\,. \tag{3.24}$$

Then (3.19) follows by (3.22), (3.23), (3.24). □

Note that the Hamiltonian $H_2 + H_3^{(3)} + H_4^{(4)}$ (see (3.19)) possesses the invariant subspace $\{z = 0\}$ and the system restricted to $\{z = 0\}$ is completely integrable and non-isochronous (actually it is formed by ν decoupled rotators). We shall construct quasi-periodic solutions which bifurcate from this invariant manifold.

In order to enter in a perturbative regime, we have to eliminate further monomials of $H^{(4)}$ in (3.19). The minimal requirement for the convergence of the nonlinear Nash–Moser iteration is to eliminate the monomials $R(v^5)$ and $R(v^4 z)$. Here we need the choice of the sites of Hypothesis ($\mathtt{S}2$).

Remark 3.5. In the KAM theorems [25,28] (and [30,32]), as well as for the perturbed mKdV equations (1.10), these further steps of Birkhoff normal form are not required because the nonlinearity of the original PDE is yet cubic. A difficulty of KdV is that the nonlinearity is quadratic.

We spell out Hypothesis ($\mathtt{S}2$) as follows:

- ($\mathtt{S}2_0$). There is no choice of 5 integers $j_1, \ldots, j_5 \in S$ such that

$$j_1 + \ldots + j_5 = 0\,, \quad j_1^3 + \ldots + j_5^3 = 0\,. \tag{3.25}$$

- ($\mathtt{S}2_1$). There is no choice of 4 integers $j_1, \ldots, j_4$ in S and an integer in the complementary set $j_5 \in S^c := (\mathbb{Z} \setminus \{0\}) \setminus S$ such that (3.25) holds.

The homogeneous component of degree 5 of $H^{(4)}$ is

$$H_5^{(4)}(u) = \sum_{j_1+\ldots+j_5=0} H^{(4)}_{j_1,\ldots,j_5} u_{j_1} \ldots u_{j_5}\,.$$

We want to remove from $H_5^{(4)}$ the terms with at most one index among $j_1, \ldots, j_5$ outside S. We consider the auxiliary Hamiltonian

$$F^{(5)} = \sum_{\substack{j_1+\ldots+j_5=0 \\ \text{at most one index outside } S}} F^{(5)}_{j_1\ldots j_5} u_{j_1} \ldots u_{j_5} , \quad F^{(5)}_{j_1\ldots j_5} := \frac{H^{(5)}_{j_1\ldots j_5}}{\mathrm{i}(j_1^3 + \ldots + j_5^3)} . \tag{3.26}$$

By Hypotheses (S2_0), (S2_1), if $j_1 + \ldots + j_5 = 0$ with at most one index outside S then $j_1^3 + \ldots + j_5^3 \neq 0$ and $F^{(5)}$ is well defined. Let $\Phi^{(5)}$ be the time 1-flow generated by $X_{F^{(5)}}$. The new Hamiltonian is

$$H^{(5)} := H^{(4)} \circ \Phi^{(5)} = H_2 + H_3^{(3)} + H_4^{(4)} + \{H_2, F^{(5)}\} + H_5^{(4)} + H_{\geq 6}^{(5)} \tag{3.27}$$

where, by (3.26),

$$H_5^{(5)} := \{H_2, F^{(5)}\} + H_5^{(4)} = \sum_{q=2}^{5} R(v^{5-q} z^q) .$$

Renaming $\mathcal{H} := H^{(5)}$, namely $\mathcal{H}_n := H_n^{(n)}$, $n = 3, 4, 5$, and setting $\Phi_B := \Phi^{(3)} \circ \Phi^{(4)} \circ \Phi^{(5)}$, formula (3.5) follows.

The homogeneous component $H_5^{(4)}$ preserves the momentum, see Section 2.1. Hence $F^{(5)}$ also preserves the momentum. As a consequence, also $H_k^{(5)}$, $k \leq 5$, preserve the momentum.

Finally, since $F^{(5)}$ is Fourier-supported on a finite set, the transformation $\Phi^{(5)}$ is of type (3.4) (and analytic), and therefore also the composition Φ_B is of type (3.4) (and analytic).

4. Action-angle variables

We now introduce action-angle variables on the tangential directions by the change of coordinates

$$\begin{cases} u_j := \sqrt{\xi_j + |j| y_j}\, e^{\mathrm{i}\theta_j} , & \text{if } j \in S , \\ u_j := z_j , & \text{if } j \in S^c , \end{cases} \tag{4.1}$$

where (recall $u_{-j} = \bar{u}_j$)

$$\xi_{-j} = \xi_j , \quad \xi_j > 0 , \quad y_{-j} = y_j , \quad \theta_{-j} = -\theta_j , \quad \theta_j, y_j \in \mathbb{R} , \quad \forall j \in S . \tag{4.2}$$

For the tangential sites $S^+ := \{\bar{\jmath}_1, \ldots, \bar{\jmath}_\nu\}$ we shall also denote $\theta_{\bar{\jmath}_i} := \theta_i$, $y_{\bar{\jmath}_i} := y_i$, $\xi_{\bar{\jmath}_i} := \xi_i$, $i = 1, \ldots \nu$.

The symplectic 2-form Ω in (2.5) (i.e. (2.1)) becomes

$$\mathcal{W} := \sum_{i=1}^{\nu} d\theta_i \wedge dy_i + \frac{1}{2} \sum_{j \in S^c \setminus \{0\}} \frac{1}{\mathrm{i} j} dz_j \wedge dz_{-j} = \Big(\sum_{i=1}^{\nu} d\theta_i \wedge dy_i \Big) \oplus \Omega_{S^\perp} = d\Lambda \tag{4.3}$$

where $\Omega_{S^\perp}$ denotes the restriction of Ω to $H_S^\perp$ (see (2.8)) and Λ is the contact 1-form on $\mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$ defined by $\Lambda_{(\theta,y,z)} : \mathbb{R}^\nu \times \mathbb{R}^\nu \times H_S^\perp \to \mathbb{R}$,

$$\Lambda_{(\theta,y,z)}[\widehat{\theta}, \widehat{y}, \widehat{z}] := -y \cdot \widehat{\theta} + \frac{1}{2} (\partial_x^{-1} z, \widehat{z})_{L^2(\mathbb{T})} . \tag{4.4}$$

Instead of working in a shrinking neighborhood of the origin, it is a convenient devise to rescale the "unperturbed actions" ξ and the action-angle variables as

$$\xi \mapsto \varepsilon^2 \xi , \quad y \mapsto \varepsilon^{2b} y , \quad z \mapsto \varepsilon^b z . \tag{4.5}$$

Then the symplectic 2-form in (4.3) transforms into $\varepsilon^{2b}\mathcal{W}$. Hence the Hamiltonian system generated by $\mathcal{H}$ in (3.5) transforms into the new Hamiltonian system

$$\dot{\theta} = \partial_y H_\varepsilon(\theta, y, z) , \ \dot{y} = -\partial_\theta H_\varepsilon(\theta, y, z) , \ z_t = \partial_x \nabla_z H_\varepsilon(\theta, y, z) , \quad H_\varepsilon := \varepsilon^{-2b} \mathcal{H} \circ A_\varepsilon \tag{4.6}$$

where

$$A_\varepsilon(\theta, y, z) := \varepsilon v_\varepsilon(\theta, y) + \varepsilon^b z := \varepsilon \sum_{j \in S} \sqrt{\xi_j + \varepsilon^{2(b-1)} |j| y_j}\, e^{\mathrm{i}\theta_j} e^{\mathrm{i} j x} + \varepsilon^b z . \tag{4.7}$$

We shall still denote by $X_{H_\varepsilon} = (\partial_y H_\varepsilon, -\partial_\theta H_\varepsilon, \partial_x \nabla_z H_\varepsilon)$ the Hamiltonian vector field in the variables $(\theta, y, z) \in \mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$.

We now write explicitly the Hamiltonian $H_\varepsilon(\theta, y, z)$ in (4.6). The quadratic Hamiltonian H_2 in (3.1) transforms into

$$\varepsilon^{-2b} H_2 \circ A_\varepsilon = const + \sum_{j \in S^+} j^3 y_j + \frac{1}{2} \int_{\mathbb{T}} z_x^2 \, dx \,, \tag{4.8}$$

and, recalling (3.6), (3.7), the Hamiltonian $\mathcal{H}$ in (3.5) transforms into (shortly writing $v_\varepsilon := v_\varepsilon(\theta, y)$)

$$\begin{aligned} H_\varepsilon(\theta, y, z) &= e(\xi) + \alpha(\xi) \cdot y + \frac{1}{2} \int_{\mathbb{T}} z_x^2 dx + \varepsilon^b \int_{\mathbb{T}} z^3 dx + 3\varepsilon \int_{\mathbb{T}} v_\varepsilon z^2 dx \\ &\quad + \varepsilon^2 \Big\{ 6 \int_{\mathbb{T}} v_\varepsilon z \Pi_S \big((\partial_x^{-1} v_\varepsilon)(\partial_x^{-1} z) \big) \, dx + 3 \int_{\mathbb{T}} z^2 \pi_0 (\partial_x^{-1} v_\varepsilon)^2 \, dx \Big\} - \frac{3}{2} \varepsilon^{2b} \sum_{j \in S} y_j^2 \\ &\quad + \varepsilon^{b+1} R(v_\varepsilon z^3) + \varepsilon^3 R(v_\varepsilon^3 z^2) + \varepsilon^{2+b} \sum_{q=3}^{5} \varepsilon^{(q-3)(b-1)} R(v_\varepsilon^{5-q} z^q) + \varepsilon^{-2b} \mathcal{H}_{\geq 6}(\varepsilon v_\varepsilon + \varepsilon^b z) \end{aligned} \tag{4.9}$$

where $e(\xi)$ is a constant, and the frequency-amplitude map is

$$\alpha(\xi) := \bar{\omega} + \varepsilon^2 \mathbb{A} \xi \,, \quad \mathbb{A} := -6 \, \mathrm{diag}\{1/j\}_{j \in S^+} \,. \tag{4.10}$$

We write the Hamiltonian in (4.9) as

$$H_\varepsilon = \mathcal{N} + P \,, \quad \mathcal{N}(\theta, y, z) = \alpha(\xi) \cdot y + \frac{1}{2} \big(N(\theta) z, z \big)_{L^2(\mathbb{T})} \,, \tag{4.11}$$

where

$$\begin{aligned} \frac{1}{2} \big(N(\theta) z, z \big)_{L^2(\mathbb{T})} &:= \frac{1}{2} \big((\partial_z \nabla H_\varepsilon)(\theta, 0, 0)[z], z \big)_{L^2(\mathbb{T})} = \frac{1}{2} \int_{\mathbb{T}} z_x^2 dx + 3\varepsilon \int_{\mathbb{T}} v_\varepsilon(\theta, 0) z^2 dx \\ &\quad + \varepsilon^2 \Big\{ 6 \int_{\mathbb{T}} v_\varepsilon(\theta, 0) z \Pi_S \big((\partial_x^{-1} v_\varepsilon(\theta, 0))(\partial_x^{-1} z) \big) dx + 3 \int_{\mathbb{T}} z^2 \pi_0 (\partial_x^{-1} v_\varepsilon(\theta, 0))^2 dx \Big\} + \ldots \end{aligned} \tag{4.12}$$

and $P := H_\varepsilon - \mathcal{N}$.

5. The nonlinear functional setting

We look for an embedded invariant torus

$$i : \mathbb{T}^\nu \to \mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp \,, \quad \varphi \mapsto i(\varphi) := (\theta(\varphi), y(\varphi), z(\varphi)) \tag{5.1}$$

of the Hamiltonian vector field X_{H_ε} filled by quasi-periodic solutions with diophantine frequency ω. We require that ω belongs to the set

$$\Omega_\varepsilon := \alpha([1, 2]^\nu) = \{ \alpha(\xi) : \xi \in [1, 2]^\nu \} \tag{5.2}$$

where α is the diffeomorphism (4.10), and, in the Hamiltonian H_ε in (4.11), we choose

$$\xi = \alpha^{-1}(\omega) = \varepsilon^{-2} \mathbb{A}^{-1} (\omega - \bar{\omega}) \,. \tag{5.3}$$

Since any $\omega \in \Omega_\varepsilon$ is ε^2-close to the integer vector $\bar{\omega}$ (see (2.10)), we require that the constant γ in the diophantine inequality

$$|\omega \cdot l| \geq \gamma \langle l \rangle^{-\tau} \,, \quad \forall l \in \mathbb{Z}^\nu \setminus \{0\} \,, \quad \text{satisfies } \gamma = \varepsilon^{2+a} \quad \text{for some } a > 0 \,. \tag{5.4}$$

We remark that the definition of γ in (5.4) is slightly stronger than the minimal condition, which is $\gamma \leq c\varepsilon^2$ with c small enough. In addition to (5.4) we shall also require that ω satisfies the first and second order Melnikov-non-resonance conditions (8.120).

We look for an embedded invariant torus of the modified Hamiltonian vector field $X_{H_{\varepsilon,\zeta}} = X_{H_\varepsilon} + (0, \zeta, 0)$ which is generated by the Hamiltonian

$$H_{\varepsilon,\zeta}(\theta, y, z) := H_\varepsilon(\theta, y, z) + \zeta \cdot \theta , \quad \zeta \in \mathbb{R}^\nu . \tag{5.5}$$

Note that $X_{H_{\varepsilon,\zeta}}$ is periodic in θ (unlike $H_{\varepsilon,\zeta}$). It turns out that an invariant torus for $X_{H_{\varepsilon,\zeta}}$ is actually invariant for X_{H_ε}, see Lemma 6.1. We introduce the parameter $\zeta \in \mathbb{R}^\nu$ in order to control the average in the y-component of the linearized equations. Thus we look for zeros of the nonlinear operator

$$\mathcal{F}(i, \zeta) := \mathcal{F}(i, \zeta, \omega, \varepsilon) := \mathcal{D}_\omega i(\varphi) - X_{H_{\varepsilon,\zeta}}(i(\varphi)) = \mathcal{D}_\omega i(\varphi) - X_{\mathcal{N}}(i(\varphi)) - X_P(i(\varphi)) + (0, \zeta, 0) \tag{5.6}$$

$$:= \begin{pmatrix} \mathcal{D}_\omega \theta(\varphi) - \partial_y H_\varepsilon(i(\varphi)) \\ \mathcal{D}_\omega y(\varphi) + \partial_\theta H_\varepsilon(i(\varphi)) + \zeta \\ \mathcal{D}_\omega z(\varphi) - \partial_x \nabla_z H_\varepsilon(i(\varphi)) \end{pmatrix} = \begin{pmatrix} \mathcal{D}_\omega \Theta(\varphi) - \partial_y P(i(\varphi)) \\ \mathcal{D}_\omega y(\varphi) + \frac{1}{2}\partial_\theta (N(\theta(\varphi))z(\varphi), z(\varphi))_{L^2(\mathbb{T})} + \partial_\theta P(i(\varphi)) + \zeta \\ \mathcal{D}_\omega z(\varphi) - \partial_x N(\theta(\varphi))z(\varphi) - \partial_x \nabla_z P(i(\varphi)) \end{pmatrix}$$

where $\Theta(\varphi) := \theta(\varphi) - \varphi$ is $(2\pi)^\nu$-periodic and we use the short notation

$$\mathcal{D}_\omega := \omega \cdot \partial_\varphi . \tag{5.7}$$

The Sobolev norm of the periodic component of the embedded torus

$$\mathfrak{I}(\varphi) := i(\varphi) - (\varphi, 0, 0) := (\Theta(\varphi), y(\varphi), z(\varphi)) , \quad \Theta(\varphi) := \theta(\varphi) - \varphi , \tag{5.8}$$

is

$$\|\mathfrak{I}\|_s := \|\Theta\|_{H^s_\varphi} + \|y\|_{H^s_\varphi} + \|z\|_s \tag{5.9}$$

where $\|z\|_s := \|z\|_{H^s_{\varphi,x}}$ is defined in (2.11). We link the rescaling (4.5) with the diophantine constant $\gamma = \varepsilon^{2+a}$ by choosing

$$\gamma = \varepsilon^{2b} , \qquad b = 1 + (a/2) . \tag{5.10}$$

Other choices are possible, see Remark 5.2.

Theorem 5.1. *Let the tangential sites S in (1.8) satisfy* (S1), (S2). *Then, for all $\varepsilon \in (0, \varepsilon_0)$, where ε_0 is small enough, there exists a Cantor-like set $\mathcal{C}_\varepsilon \subset \Omega_\varepsilon$, with asymptotically full measure as $\varepsilon \to 0$, namely*

$$\lim_{\varepsilon \to 0} \frac{|\mathcal{C}_\varepsilon|}{|\Omega_\varepsilon|} = 1 , \tag{5.11}$$

such that, for all $\omega \in \mathcal{C}_\varepsilon$, there exists a solution $i_\infty(\varphi) := i_\infty(\omega, \varepsilon)(\varphi)$ of $\mathcal{D}_\omega i_\infty(\varphi) - X_{H_\varepsilon}(i_\infty(\varphi)) = 0$. Hence the embedded torus $\varphi \mapsto i_\infty(\varphi)$ is invariant for the Hamiltonian vector field $X_{H_\varepsilon(\cdot,\xi)}$ with ξ as in (5.3), and it is filled by quasi-periodic solutions with frequency ω. The torus i_∞ satisfies

$$\|i_\infty(\varphi) - (\varphi, 0, 0)\|^{\mathrm{Lip}(\gamma)}_{s_0+\mu} = O(\varepsilon^{6-2b}\gamma^{-1}) \tag{5.12}$$

for some $\mu := \mu(\nu) > 0$. Moreover, the torus i_∞ is LINEARLY STABLE.

Theorem 5.1 is proved in Sections 6–9. It implies Theorem 1.1 where the ξ_j in (1.9) are $\varepsilon^2\xi_j$, $\xi_j \in [1, 2]$, in (5.3). By (5.12), going back to the variables before the rescaling (4.5), we get $\Theta_\infty = O(\varepsilon^{6-2b}\gamma^{-1})$, $y_\infty = O(\varepsilon^6\gamma^{-1})$, $z_\infty = O(\varepsilon^{6-b}\gamma^{-1})$, which, as $b \to 1^+$, tend to the expected optimal estimates.

Remark 5.2. There are other possible ways to link the rescaling (4.5) with the diophantine constant $\gamma = \varepsilon^{2+a}$. The choice $\gamma > \varepsilon^{2b}$ reduces to study perturbations of an isochronous system (as in [23,25,28]), and it is convenient to introduce $\xi(\omega)$ as a variable. The case $\varepsilon^{2b} > \gamma$, in particular $b = 1$, has to be dealt with a perturbation approach of a non-isochronous system à la Arnold–Kolmogorov.

We now give the tame estimates for the composition operator induced by the Hamiltonian vector fields $X_{\mathcal{N}}$ and X_P in (5.6), that we shall use in the next sections.

We first estimate the composition operator induced by $v_\varepsilon(\theta, y)$ defined in (4.7). Since the functions $y \mapsto \sqrt{\xi + \varepsilon^{2(b-1)}|j|y}$, $\theta \mapsto e^{i\theta}$ are analytic for ε small enough and $|y| \le C$, the composition Lemma 2.2 implies that, for all $\Theta, y \in H^s(\mathbb{T}^\nu, \mathbb{R}^\nu)$, $\|\Theta\|_{s_0}, \|y\|_{s_0} \le 1$, setting $\theta(\varphi) := \varphi + \Theta(\varphi)$, $\|v_\varepsilon(\theta(\varphi), y(\varphi))\|_s \le_s 1 + \|\Theta\|_s + \|y\|_s$. Hence, using also (5.3), the map A_ε in (4.7) satisfies, for all $\|\mathfrak{I}\|_{s_0}^{\mathrm{Lip}(\gamma)} \le 1$ (see (5.8))

$$\|A_\varepsilon(\theta(\varphi), y(\varphi), z(\varphi))\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon(1 + \|\mathfrak{I}\|_s^{\mathrm{Lip}(\gamma)}) . \tag{5.13}$$

We now give tame estimates for the Hamiltonian vector fields $X_{\mathcal{N}}$, X_P, X_{H_ε}, see (4.11)–(4.12).

Lemma 5.3. *Let $\mathfrak{I}(\varphi)$ in (5.8) satisfy $\|\mathfrak{I}\|_{s_0+3}^{\mathrm{Lip}(\gamma)} \le C\varepsilon^{6-2b}\gamma^{-1}$. Then*

$$\|\partial_y P(i)\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^4 + \varepsilon^{2b}\|\mathfrak{I}\|_{s+1}^{\mathrm{Lip}(\gamma)} , \qquad \|\partial_\theta P(i)\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^{6-2b}(1 + \|\mathfrak{I}\|_{s+1}^{\mathrm{Lip}(\gamma)}) \tag{5.14}$$

$$\|\nabla_z P(i)\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^{5-b} + \varepsilon^{6-b}\gamma^{-1}\|\mathfrak{I}\|_{s+1}^{\mathrm{Lip}(\gamma)} , \qquad \|X_P(i)\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^{6-2b} + \varepsilon^{2b}\|\mathfrak{I}\|_{s+3}^{\mathrm{Lip}(\gamma)} \tag{5.15}$$

$$\|\partial_\theta\partial_y P(i)\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^4 + \varepsilon^5\gamma^{-1}\|\mathfrak{I}\|_{s+2}^{\mathrm{Lip}(\gamma)} , \qquad \|\partial_y\nabla_z P(i)\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^{b+3} + \varepsilon^{2b-1}\|\mathfrak{I}\|_{s+2}^{\mathrm{Lip}(\gamma)} \tag{5.16}$$

$$\|\partial_{yy} P(i) + 3\varepsilon^{2b} I_{\mathbb{R}^\nu}\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^{2+2b} + \varepsilon^{2b+3}\gamma^{-1}\|\mathfrak{I}\|_{s+2}^{\mathrm{Lip}(\gamma)} \tag{5.17}$$

and, for all $\widehat{\imath} := (\widehat{\Theta}, \widehat{y}, \widehat{z})$,

$$\|\partial_y d_i X_P(i)[\widehat{\imath}]\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^{2b-1}\big(\|\widehat{\imath}\|_{s+3}^{\mathrm{Lip}(\gamma)} + \|\mathfrak{I}\|_{s+3}^{\mathrm{Lip}(\gamma)}\|\widehat{\imath}\|_{s_0+3}^{\mathrm{Lip}(\gamma)}\big) \tag{5.18}$$

$$\|d_i X_{H_\varepsilon}(i)[\widehat{\imath}] + (0, 0, \partial_{xxx}\hat{z})\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon\big(\|\widehat{\imath}\|_{s+3}^{\mathrm{Lip}(\gamma)} + \|\mathfrak{I}\|_{s+3}^{\mathrm{Lip}(\gamma)}\|\widehat{\imath}\|_{s_0+3}^{\mathrm{Lip}(\gamma)}\big) \tag{5.19}$$

$$\|d_i^2 X_{H_\varepsilon}(i)[\widehat{\imath}, \widehat{\imath}]\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon\Big(\|\widehat{\imath}\|_{s+3}^{\mathrm{Lip}(\gamma)}\|\widehat{\imath}\|_{s_0+3}^{\mathrm{Lip}(\gamma)} + \|\mathfrak{I}\|_{s+3}^{\mathrm{Lip}(\gamma)}(\|\widehat{\imath}\|_{s_0+3}^{\mathrm{Lip}(\gamma)})^2\Big). \tag{5.20}$$

In the sequel we will also use that, by the diophantine condition (5.4), the operator $\mathcal{D}_\omega^{-1}$ (see (5.7)) is defined for all functions u with zero φ-average, and satisfies

$$\|\mathcal{D}_\omega^{-1}u\|_s \le C\gamma^{-1}\|u\|_{s+\tau} , \quad \|\mathcal{D}_\omega^{-1}u\|_s^{\mathrm{Lip}(\gamma)} \le C\gamma^{-1}\|u\|_{s+2\tau+1}^{\mathrm{Lip}(\gamma)} . \tag{5.21}$$

6. Approximate inverse

In order to implement a convergent Nash–Moser scheme that leads to a solution of $\mathcal{F}(i, \zeta) = 0$ our aim is to construct an *approximate right inverse* (which satisfies tame estimates) of the linearized operator

$$d_{i,\zeta}\mathcal{F}(i_0, \zeta_0)[\widehat{\imath}, \widehat{\zeta}] = d_{i,\zeta}\mathcal{F}(i_0)[\widehat{\imath}, \widehat{\zeta}] = \mathcal{D}_\omega\widehat{\imath} - d_i X_{H_\varepsilon}(i_0(\varphi))[\widehat{\imath}] + (0, \widehat{\zeta}, 0) , \tag{6.1}$$

see Theorem 6.10. Note that $d_{i,\zeta}\mathcal{F}(i_0, \zeta_0) = d_{i,\zeta}\mathcal{F}(i_0)$ is independent of ζ_0 (see (5.6)).

The notion of approximate right inverse is introduced in [35]. It denotes a linear operator which is an *exact* right inverse at a solution (i_0, ζ_0) of $\mathcal{F}(i_0, \zeta_0) = 0$. We want to implement the general strategy in [7,8] which reduces the search of an approximate right inverse of (6.1) to the search of an approximate inverse on the normal directions only.

It is well known that an invariant torus i_0 with diophantine flow is isotropic (see e.g. [7]), namely the pull-back 1-form $i_0^*\Lambda$ is closed, where Λ is the contact 1-form in (4.4). This is tantamount to say that the 2-form $\mathcal{W}$ (see (4.3)) vanishes on the torus $i_0(\mathbb{T}^\nu)$ (i.e. $\mathcal{W}$ vanishes on the tangent space at each point $i_0(\varphi)$ of the manifold $i_0(\mathbb{T}^\nu)$), because $i_0^*\mathcal{W} = i_0^* d\Lambda = d i_0^*\Lambda$. For an "approximately invariant" torus i_0 the 1-form $i_0^*\Lambda$ is only "approximately closed". In order to make this statement quantitative we consider

$$i_0^*\Lambda = \sum_{k=1}^{\nu} a_k(\varphi) d\varphi_k , \quad a_k(\varphi) := -\big([\partial_\varphi\theta_0(\varphi)]^T y_0(\varphi)\big)_k + \frac{1}{2}(\partial_{\varphi_k} z_0(\varphi), \partial_x^{-1} z_0(\varphi))_{L^2(\mathbb{T})} \tag{6.2}$$

and we quantify how small is

$$i_0^*\mathcal{W} = d\, i_0^*\Lambda = \sum_{1\le k<j\le \nu} A_{kj}(\varphi) d\varphi_k \wedge d\varphi_j\,, \quad A_{kj}(\varphi) := \partial_{\varphi_k} a_j(\varphi) - \partial_{\varphi_j} a_k(\varphi)\,. \tag{6.3}$$

Along this section we will always assume the following hypothesis (which will be verified at each step of the Nash–Moser iteration):

- ASSUMPTION. The map $\omega \mapsto i_0(\omega)$ is a Lipschitz function defined on some subset $\Omega_o \subset \Omega_\varepsilon$, where Ω_ε is defined in (5.2), and, for some $\mu := \mu(\tau,\nu) > 0$,
$$\|\mathfrak{I}_0\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)} \le C\varepsilon^{6-2b}\gamma^{-1}, \quad \|Z\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)} \le C\varepsilon^{6-2b}, \quad \gamma = \varepsilon^{2+a}, \quad b := 1 + (a/2)\,, \quad a \in (0, 1/6), \tag{6.4}$$
where $\mathfrak{I}_0(\varphi) := i_0(\varphi) - (\varphi, 0, 0)$, and
$$Z(\varphi) := (Z_1, Z_2, Z_3)(\varphi) := \mathcal{F}(i_0, \zeta_0)(\varphi) = \omega \cdot \partial_\varphi i_0(\varphi) - X_{H_{\varepsilon,\zeta_0}}(i_0(\varphi))\,. \tag{6.5}$$

Lemma 6.1. $|\zeta_0|^{\mathrm{Lip}(\gamma)} \le C\|Z\|_{s_0}^{\mathrm{Lip}(\gamma)}$. *If $\mathcal{F}(i_0,\zeta_0) = 0$ then $\zeta_0 = 0$, namely the torus i_0 is invariant for X_{H_ε}.*

Proof. It is proved in [7] the formula

$$\zeta_0 = \int_{\mathbb{T}^\nu} -[\partial_\varphi y_0(\varphi)]^T Z_1(\varphi) + [\partial_\varphi \theta_0(\varphi)]^T Z_2(\varphi) - [\partial_\varphi z_0(\varphi)]^T \partial_x^{-1} Z_3(\varphi)\, d\varphi\,.$$

Hence the lemma follows by (6.4) and usual algebra estimate. □

We now quantify the size of $i_0^*\mathcal{W}$ in terms of Z. Directly from (6.2) and (6.3) one has $\|A_{kj}\|_s^{\mathrm{Lip}(\gamma)} \le_s \|\mathfrak{I}_0\|_{s+2}^{\mathrm{Lip}(\gamma)}$. Moreover, A_{kj} also satisfies the following bound.

Lemma 6.2. *The coefficients $A_{kj}(\varphi)$ in (6.3) satisfy*

$$\|A_{kj}\|_s^{\mathrm{Lip}(\gamma)} \le_s \gamma^{-1}\big(\|Z\|_{s+2\tau+2}^{\mathrm{Lip}(\gamma)} + \|Z\|_{s_0+1}^{\mathrm{Lip}(\gamma)}\|\mathfrak{I}_0\|_{s+2\tau+2}^{\mathrm{Lip}(\gamma)}\big)\,. \tag{6.6}$$

Proof. We estimate the coefficients of the Lie derivative $L_\omega(i_0^*\mathcal{W}) := \sum_{k<j}\mathcal{D}_\omega A_{kj}(\varphi) d\varphi_k \wedge d\varphi_j$. Denoting by $\underline{e}_k$ the k-th versor of $\mathbb{R}^\nu$ we have

$$\mathcal{D}_\omega A_{kj} = L_\omega(i_0^*\mathcal{W})(\varphi)[\underline{e}_k, \underline{e}_j] = \mathcal{W}\big(\partial_\varphi Z(\varphi)\underline{e}_k, \partial_\varphi i_0(\varphi)\underline{e}_j\big) + \mathcal{W}\big(\partial_\varphi i_0(\varphi)\underline{e}_k, \partial_\varphi Z(\varphi)\underline{e}_j\big)$$

(see [7]). Hence

$$\|\mathcal{D}_\omega A_{kj}\|_s^{\mathrm{Lip}(\gamma)} \le_s \|Z\|_{s+1}^{\mathrm{Lip}(\gamma)} + \|Z\|_{s_0+1}^{\mathrm{Lip}(\gamma)}\|\mathfrak{I}_0\|_{s+1}^{\mathrm{Lip}(\gamma)}\,. \tag{6.7}$$

The bound (6.6) follows applying $\mathcal{D}_\omega^{-1}$ and using (5.21). □

As in [7] we first modify the approximate torus i_0 to obtain an isotropic torus i_δ which is still approximately invariant. We denote the Laplacian $\Delta_\varphi := \sum_{k=1}^\nu \partial_{\varphi_k}^2$.

Lemma 6.3 *(Isotropic torus). The torus $i_\delta(\varphi) := (\theta_0(\varphi), y_\delta(\varphi), z_0(\varphi))$ defined by*

$$y_\delta := y_0 + [\partial_\varphi\theta_0(\varphi)]^{-T}\rho(\varphi)\,, \qquad \rho_j(\varphi) := \Delta_\varphi^{-1}\sum_{k=1}^\nu \partial_{\varphi_j} A_{kj}(\varphi) \tag{6.8}$$

is isotropic. If (6.4) holds, then, for some $\sigma := \sigma(\nu,\tau)$,

$$\|y_\delta - y_0\|_s^{\mathrm{Lip}(\gamma)} \le_s \|\mathfrak{I}_0\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}\,, \tag{6.9}$$

$$\|y_\delta - y_0\|_s^{\mathrm{Lip}(\gamma)} \le_s \gamma^{-1}\big\{\|Z\|_{s+\sigma}^{\mathrm{Lip}(\gamma)} + \|Z\|_{s_0+\sigma}^{\mathrm{Lip}(\gamma)}\|\mathfrak{I}_0\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}\big\}\,, \tag{6.10}$$

$$\|\mathcal{F}(i_\delta,\zeta_0)\|_s^{\mathrm{Lip}(\gamma)} \le_s \|Z\|_{s+\sigma}^{\mathrm{Lip}(\gamma)} + \varepsilon^{2b-1}\gamma^{-1}\|\mathfrak{I}_0\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}\|Z\|_{s_0+\sigma}^{\mathrm{Lip}(\gamma)}\,, \tag{6.11}$$

$$\|\partial_i[i_\delta][\hat\imath]\|_s \le_s \|\hat\imath\|_s + \|\mathfrak{I}_0\|_{s+\sigma}\|\hat\imath\|_s\,. \tag{6.12}$$

In the paper we denote equivalently the differential by ∂_i or d_i. Moreover we denote by $\sigma := \sigma(\nu, \tau)$ possibly different (larger) "loss of derivatives" constants.

Proof. In this proof we write $\| \ \|_s$ to denote $\| \ \|_s^{\mathrm{Lip}(\gamma)}$. The proof of the isotropy of i_δ is in [7]. The estimate (6.9) follows by (6.8), (6.2), (6.3), (6.4). The estimate (6.10) follows by (6.8), (6.6), (6.4) and the tame bound for the inverse $\|[\partial_\varphi \theta_0]^{-T}\|_s \leq_s 1 + \|\mathfrak{I}_0\|_{s+1}$. It remains to estimate the difference (see (5.6) and note that $X_{\mathcal{N}}$ does not depend on y)

$$\mathcal{F}(i_\delta, \zeta_0) - \mathcal{F}(i_0, \zeta_0) = \begin{pmatrix} 0 \\ \mathcal{D}_\omega(y_\delta - y_0) \\ 0 \end{pmatrix} + X_P(i_\delta) - X_P(i_0) \,. \tag{6.13}$$

Using (5.16), (5.17), we get $\|\partial_y X_P(i)\|_s \leq_s \varepsilon^{2b} + \varepsilon^{2b-1}\|\mathfrak{I}\|_{s+3}$. Hence (6.9), (6.10), (6.4) imply

$$\|X_P(i_\delta) - X_P(i_0)\|_s \leq_s \|Z\|_{s+\sigma} + \varepsilon^{2b-1}\gamma^{-1}\|\mathfrak{I}_0\|_{s+\sigma}\|Z\|_{s_0+\sigma} \,. \tag{6.14}$$

Differentiating (6.8) we have

$$\mathcal{D}_\omega(y_\delta - y_0) = [\partial_\varphi \theta_0(\varphi)]^{-T}\mathcal{D}_\omega \rho(\varphi) + (\mathcal{D}_\omega[\partial_\varphi \theta_0(\varphi)]^{-T})\rho(\varphi) \tag{6.15}$$

and $\mathcal{D}_\omega \rho_j(\varphi) = \Delta_\varphi^{-1}\sum_{k=1}^{\nu}\partial_{\varphi_j}\mathcal{D}_\omega A_{kj}(\varphi)$. Using (6.7), we deduce that

$$\|[\partial_\varphi \theta_0]^{-T}\mathcal{D}_\omega \rho\|_s \leq_s \|Z\|_{s+1} + \|Z\|_{s_0+1}\|\mathfrak{I}_0\|_{s+1} \,. \tag{6.16}$$

To estimate the second term in (6.15), we differentiate $Z_1(\varphi) = \mathcal{D}_\omega \theta_0(\varphi) - \omega - (\partial_y P)(i_0(\varphi))$ (which is the first component in (5.6)) with respect to φ. We get $\mathcal{D}_\omega \partial_\varphi \theta_0(\varphi) = \partial_\varphi(\partial_y P)(i_0(\varphi)) + \partial_\varphi Z_1(\varphi)$. Then, by (5.14),

$$\|\mathcal{D}_\omega[\partial_\varphi \theta_0]^T\|_s \leq_s \varepsilon^4 + \varepsilon^{2b}\|\mathfrak{I}_0\|_{s+2} + \|Z\|_{s+1} \,. \tag{6.17}$$

Since $\mathcal{D}_\omega[\partial_\varphi \theta_0(\varphi)]^{-T} = -[\partial_\varphi \theta_0(\varphi)]^{-T}\big(\mathcal{D}_\omega[\partial_\varphi \theta_0(\varphi)]^T\big)[\partial_\varphi \theta_0(\varphi)]^{-T}$, the bounds (6.17), (6.6), (6.4) imply

$$\|(\mathcal{D}_\omega[\partial_\varphi \theta_0]^{-T})\rho\|_s \leq_s \varepsilon^{6-2b}\gamma^{-1}\|Z\|_{s+\sigma} + \|\mathfrak{I}_0\|_{s+\sigma}\|Z\|_{s_0+\sigma} \,. \tag{6.18}$$

In conclusion (6.13), (6.14), (6.15), (6.16), (6.18) imply (6.11). The bound (6.12) follows by (6.8), (6.3), (6.2), (6.4). □

In order to find an approximate inverse of the linearized operator $d_{i,\zeta}\mathcal{F}(i_\delta)$ we introduce a suitable set of symplectic coordinates nearby the isotropic torus i_δ. We consider the map $G_\delta : (\psi, \eta, w) \to (\theta, y, z)$ of the phase space $\mathbb{T}^\nu \times \mathbb{R}^\nu \times H_S^\perp$ defined by

$$\begin{pmatrix} \theta \\ y \\ z \end{pmatrix} := G_\delta \begin{pmatrix} \psi \\ \eta \\ w \end{pmatrix} := \begin{pmatrix} \theta_0(\psi) \\ y_\delta(\psi) + [\partial_\psi \theta_0(\psi)]^{-T}\eta + \big[(\partial_\theta \tilde{z}_0)(\theta_0(\psi))\big]^T \partial_x^{-1} w \\ z_0(\psi) + w \end{pmatrix} \tag{6.19}$$

where $\tilde{z}_0(\theta) := z_0(\theta_0^{-1}(\theta))$. It is proved in [7] that G_δ is symplectic, using that the torus i_δ is isotropic (Lemma 6.3). In the new coordinates, i_δ is the trivial embedded torus $(\psi, \eta, w) = (\psi, 0, 0)$. The transformed Hamiltonian $K := K(\psi, \eta, w, \zeta_0)$ is (recall (5.5))

$$\begin{aligned} K := H_{\varepsilon,\zeta_0} \circ G_\delta = {} & \theta_0(\psi)\cdot\zeta_0 + K_{00}(\psi) + K_{10}(\psi)\cdot\eta + (K_{01}(\psi), w)_{L^2(\mathbb{T})} + \frac{1}{2}K_{20}(\psi)\eta\cdot\eta \\ & + \big(K_{11}(\psi)\eta, w\big)_{L^2(\mathbb{T})} + \frac{1}{2}\big(K_{02}(\psi)w, w\big)_{L^2(\mathbb{T})} + K_{\geq 3}(\psi, \eta, w) \end{aligned} \tag{6.20}$$

where $K_{\geq 3}$ collects the terms at least cubic in the variables (η, w). At any fixed ψ, the Taylor coefficient $K_{00}(\psi) \in \mathbb{R}$, $K_{10}(\psi) \in \mathbb{R}^\nu$, $K_{01}(\psi) \in H_S^\perp$ (it is a function of $x \in \mathbb{T}$), $K_{20}(\psi)$ is a $\nu \times \nu$ real matrix, $K_{02}(\psi)$ is a linear self-adjoint operator of $H_S^\perp$ and $K_{11}(\psi) : \mathbb{R}^\nu \to H_S^\perp$. Note that the above Taylor coefficients do not depend on the parameter ζ_0.

The Hamilton equations associated to (6.20) are

$$\begin{cases} \dot\psi = K_{10}(\psi) + K_{20}(\psi)\eta + K_{11}^T(\psi)w + \partial_\eta K_{\geq 3}(\psi,\eta,w) \\ \dot\eta = -[\partial_\psi\theta_0(\psi)]^T\zeta_0 - \partial_\psi K_{00}(\psi) - [\partial_\psi K_{10}(\psi)]^T\eta - [\partial_\psi K_{01}(\psi)]^T w \\ \qquad - \partial_\psi\big(\tfrac12 K_{20}(\psi)\eta\cdot\eta + (K_{11}(\psi)\eta, w)_{L^2(\mathbb{T})} + \tfrac12 (K_{02}(\psi)w,w)_{L^2(\mathbb{T})} + K_{\geq 3}(\psi,\eta,w)\big) \\ \dot w = \partial_x\big(K_{01}(\psi) + K_{11}(\psi)\eta + K_{02}(\psi)w + \nabla_w K_{\geq 3}(\psi,\eta,w)\big) \end{cases} \tag{6.21}$$

where $[\partial_\psi K_{10}(\psi)]^T$ is the $\nu\times\nu$ transposed matrix and $[\partial_\psi K_{01}(\psi)]^T$, $K_{11}^T(\psi) : H_S^\perp \to \mathbb{R}^\nu$ are defined by the duality relation $(\partial_\psi K_{01}(\psi)[\hat\psi], w)_{L^2} = \hat\psi\cdot[\partial_\psi K_{01}(\psi)]^T w$, $\forall\hat\psi\in\mathbb{R}^\nu$, $w\in H_S^\perp$, and similarly for K_{11}. Explicitly, for all $w\in H_S^\perp$, and denoting $\underline{e}_k$ the k-th versor of $\mathbb{R}^\nu$,

$$K_{11}^T(\psi)w = \sum_{k=1}^{\nu}\big(K_{11}^T(\psi)w\cdot\underline{e}_k\big)\underline{e}_k = \sum_{k=1}^{\nu}\big(w, K_{11}(\psi)\underline{e}_k\big)_{L^2(\mathbb{T})}\underline{e}_k \in\mathbb{R}^\nu. \tag{6.22}$$

In the next lemma we estimate the coefficients K_{00}, K_{10}, K_{01} in the Taylor expansion (6.20). Note that on an exact solution we have $Z=0$ and therefore $K_{00}(\psi)=$ const, $K_{10}=\omega$ and $K_{01}=0$.

Lemma 6.4. *Assume (6.4). Then there is $\sigma := \sigma(\tau,\nu)$ such that*

$$\|\partial_\psi K_{00}\|_s^{\mathrm{Lip}(\gamma)} + \|K_{10}-\omega\|_s^{\mathrm{Lip}(\gamma)} + \|K_{01}\|_s^{\mathrm{Lip}(\gamma)} \leq_s \|Z\|_{s+\sigma}^{\mathrm{Lip}(\gamma)} + \varepsilon^{2b-1}\gamma^{-1}\|Z\|_{s_0+\sigma}^{\mathrm{Lip}(\gamma)}\|\mathfrak{I}_0\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}.$$

Proof. Let $\mathcal{F}(i_\delta,\zeta_0) := Z_\delta := (Z_{1,\delta}, Z_{2,\delta}, Z_{3,\delta})$. By a direct calculation as in [7] (using (6.20), (5.6))

$$\begin{aligned} \partial_\psi K_{00}(\psi) &= -[\partial_\psi\theta_0(\psi)]^T\big(\zeta_0 - Z_{2,\delta} - [\partial_\psi y_\delta][\partial_\psi\theta_0]^{-1}Z_{1,\delta} + [(\partial_\theta\tilde z_0)(\theta_0(\psi))]^T\partial_x^{-1}Z_{3,\delta} \\ &\qquad + [(\partial_\theta\tilde z_0)(\theta_0(\psi))]^T\partial_x^{-1}\partial_\psi z_0(\psi)[\partial_\psi\theta_0(\psi)]^{-1}Z_{1,\delta}\big), \\ K_{10}(\psi) &= \omega - [\partial_\psi\theta_0(\psi)]^{-1}Z_{1,\delta}(\psi), \\ K_{01}(\psi) &= -\partial_x^{-1}Z_{3,\delta} + \partial_x^{-1}\partial_\psi z_0(\psi)[\partial_\psi\theta_0(\psi)]^{-1}Z_{1,\delta}(\psi). \end{aligned}$$

Then (6.4), (6.10), (6.11) and Lemma 6.1 (use also Lemma 2.4) imply the lemma. □

Remark 6.5. If $\mathcal{F}(i_0,\zeta_0)=0$ then $\zeta_0=0$ by Lemma 6.1, and Lemma 6.4 implies that (6.20) simplifies to $K = const + \omega\cdot\eta + \frac12 K_{20}(\psi)\eta\cdot\eta + \big(K_{11}(\psi)\eta, w\big)_{L^2(\mathbb{T})} + \frac12\big(K_{02}(\psi)w,w\big)_{L^2(\mathbb{T})} + K_{\geq 3}$.

We now estimate K_{20}, K_{11} in (6.20). The norm of K_{20} is the sum of the norms of its matrix entries.

Lemma 6.6. *Assume (6.4). Then*

$$\|K_{20}+3\varepsilon^{2b}I\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^{2b+2} + \varepsilon^{2b}\|\mathfrak{I}_0\|_{s+\sigma}^{\mathrm{Lip}(\gamma)} \tag{6.23}$$

$$\|K_{11}\eta\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^5\gamma^{-1}\|\eta\|_s^{\mathrm{Lip}(\gamma)} + \varepsilon^{2b-1}\|\mathfrak{I}_0\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}\|\eta\|_{s_0}^{\mathrm{Lip}(\gamma)} \tag{6.24}$$

$$\|K_{11}^T w\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^5\gamma^{-1}\|w\|_{s+2}^{\mathrm{Lip}(\gamma)} + \varepsilon^{2b-1}\|\mathfrak{I}_0\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}\|w\|_{s_0+2}^{\mathrm{Lip}(\gamma)}. \tag{6.25}$$

In particular $\|K_{20}+3\varepsilon^{2b}I\|_{s_0}^{\mathrm{Lip}(\gamma)} \leq C\varepsilon^6\gamma^{-1}$, *and*

$$\|K_{11}\eta\|_{s_0}^{\mathrm{Lip}(\gamma)} \leq C\varepsilon^5\gamma^{-1}\|\eta\|_{s_0}^{\mathrm{Lip}(\gamma)}, \qquad \|K_{11}^T w\|_{s_0}^{\mathrm{Lip}(\gamma)} \leq C\varepsilon^5\gamma^{-1}\|w\|_{s_0}^{\mathrm{Lip}(\gamma)}.$$

Proof. To shorten the notation, in this proof we write $\|\ \|_s$ for $\|\ \|_s^{\mathrm{Lip}(\gamma)}$. We have

$$K_{20}(\varphi) = [\partial_\varphi\theta_0(\varphi)]^{-1}\partial_{yy}H_\varepsilon(i_\delta(\varphi))[\partial_\varphi\theta_0(\varphi)]^{-T} = [\partial_\varphi\theta_0(\varphi)]^{-1}\partial_{yy}P(i_\delta(\varphi))[\partial_\varphi\theta_0(\varphi)]^{-T}.$$

Then (5.17), (6.4), (6.9) imply (6.23). Now (see also [7])

$$\begin{aligned} K_{11}(\varphi) &= \partial_y\nabla_z H_\varepsilon(i_\delta(\varphi))[\partial_\varphi\theta_0(\varphi)]^{-T} - \partial_x^{-1}(\partial_\theta\tilde z_0)(\theta_0(\varphi))(\partial_{yy}H_\varepsilon)(i_\delta(\varphi))[\partial_\varphi\theta_0(\varphi)]^{-T} \\ &\overset{(4.11)}{=} \partial_y\nabla_z P(i_\delta(\varphi))[\partial_\varphi\theta_0(\varphi)]^{-T} - \partial_x^{-1}(\partial_\theta\tilde z_0)(\theta_0(\varphi))(\partial_{yy}P)(i_\delta(\varphi))[\partial_\varphi\theta_0(\varphi)]^{-T}, \end{aligned}$$

therefore, using (5.16), (5.17), (6.4), (6.9), we deduce (6.24). The bound (6.25) for K_{11}^T follows by (6.22) and (6.24). $\square$

Under the linear change of variables

$$DG_\delta(\varphi,0,0)\begin{pmatrix}\widehat{\psi}\\ \widehat{\eta}\\ \widehat{w}\end{pmatrix} := \begin{pmatrix}\partial_\psi\theta_0(\varphi) & 0 & 0\\ \partial_\psi y_\delta(\varphi) & [\partial_\psi\theta_0(\varphi)]^{-T} & -[(\partial_\theta\tilde{z}_0)(\theta_0(\varphi))]^T\partial_x^{-1}\\ \partial_\psi z_0(\varphi) & 0 & I\end{pmatrix}\begin{pmatrix}\widehat{\psi}\\ \widehat{\eta}\\ \widehat{w}\end{pmatrix} \tag{6.26}$$

the linearized operator $d_{i,\zeta}\mathcal{F}(i_\delta)$ transforms (approximately, see (6.46)) into the operator obtained linearizing (6.21) at $(\psi,\eta,w,\zeta)=(\varphi,0,0,\zeta_0)$ (with $\partial_t \rightsquigarrow \mathcal{D}_\omega$), namely

$$\begin{pmatrix}\mathcal{D}_\omega\widehat{\psi} - \partial_\psi K_{10}(\varphi)[\widehat{\psi}] - K_{20}(\varphi)\widehat{\eta} - K_{11}^T(\varphi)\widehat{w}\\ \mathcal{D}_\omega\widehat{\eta} + [\partial_\psi\theta_0(\varphi)]^T\widehat{\zeta} + \partial_\psi[\partial_\psi\theta_0(\varphi)]^T[\widehat{\psi},\zeta_0] + \partial_{\psi\psi}K_{00}(\varphi)[\widehat{\psi}] + [\partial_\psi K_{10}(\varphi)]^T\widehat{\eta} + [\partial_\psi K_{01}(\varphi)]^T\widehat{w}\\ \mathcal{D}_\omega\widehat{w} - \partial_x\{\partial_\psi K_{01}(\varphi)[\widehat{\psi}] + K_{11}(\varphi)\widehat{\eta} + K_{02}(\varphi)\widehat{w}\}\end{pmatrix}. \tag{6.27}$$

We now estimate the induced composition operator.

Lemma 6.7. *Assume (6.4) and let* $\widehat{\imath} := (\widehat{\psi},\widehat{\eta},\widehat{w})$. *Then*

$$\begin{aligned}&\|DG_\delta(\varphi,0,0)[\widehat{\imath}]\|_s + \|DG_\delta(\varphi,0,0)^{-1}[\widehat{\imath}]\|_s \leq_s \|\widehat{\imath}\|_s + \|\mathfrak{I}_0\|_{s+\sigma}\|\widehat{\imath}\|_{s_0},\\ &\|D^2G_\delta(\varphi,0,0)[\widehat{\imath}_1,\widehat{\imath}_2]\|_s \leq_s \|\widehat{\imath}_1\|_s\|\widehat{\imath}_2\|_{s_0} + \|\widehat{\imath}_1\|_{s_0}\|\widehat{\imath}_2\|_s + \|\mathfrak{I}_0\|_{s+\sigma}\|\widehat{\imath}_1\|_{s_0}\|\widehat{\imath}_2\|_{s_0}\end{aligned} \tag{6.28}$$

for some $\sigma := \sigma(\nu,\tau)$. *Moreover the same estimates hold if we replace the norm* $\|\ \|_s$ *with* $\|\ \|_s^{\mathrm{Lip}(\gamma)}$.

Proof. The estimate (6.28) for $DG_\delta(\varphi,0,0)$ follows by (6.26) and (6.9). By (6.4), $\|(DG_\delta(\varphi,0,0)-I)\widehat{\imath}\|_{s_0} \leq C\varepsilon^{6-2b}\gamma^{-1}\|\widehat{\imath}\|_{s_0} \leq \|\widehat{\imath}\|_{s_0}/2$. Therefore $DG_\delta(\varphi,0,0)$ is invertible and, by Neumann series, the inverse satisfies (6.28). The bound for D^2G_δ follows by differentiating DG_δ. $\square$

In order to construct an approximate inverse of (6.27) it is sufficient to solve the equation

$$\mathbb{D}[\widehat{\psi},\widehat{\eta},\widehat{w},\widehat{\zeta}] := \begin{pmatrix}\mathcal{D}_\omega\widehat{\psi} - K_{20}(\varphi)\widehat{\eta} - K_{11}^T(\varphi)\widehat{w}\\ \mathcal{D}_\omega\widehat{\eta} + [\partial_\psi\theta_0(\varphi)]^T\widehat{\zeta}\\ \mathcal{D}_\omega\widehat{w} - \partial_x K_{11}(\varphi)\widehat{\eta} - \partial_x K_{02}(\varphi)\widehat{w}\end{pmatrix} = \begin{pmatrix}g_1\\ g_2\\ g_3\end{pmatrix} \tag{6.29}$$

which is obtained by neglecting in (6.27) the terms $\partial_\psi K_{10}$, $\partial_{\psi\psi}K_{00}$, $\partial_\psi K_{00}$, $\partial_\psi K_{01}$ and $\partial_\psi[\partial_\psi\theta_0(\varphi)]^T[\cdot,\zeta_0]$ (which are naught at a solution by Lemmata 6.4 and 6.1).

First we solve the second equation in (6.29), namely $\mathcal{D}_\omega\widehat{\eta} = g_2 - [\partial_\psi\theta_0(\varphi)]^T\widehat{\zeta}$. We choose $\widehat{\zeta}$ so that the φ-average of the right hand side is zero, namely

$$\widehat{\zeta} = \langle g_2\rangle \tag{6.30}$$

(we denote $\langle g\rangle := (2\pi)^{-\nu}\int_{\mathbb{T}^\nu} g(\varphi)d\varphi$). Note that the φ-averaged matrix $\langle[\partial_\psi\theta_0]^T\rangle = \langle I + [\partial_\psi\Theta_0]^T\rangle = I$ because $\theta_0(\varphi) = \varphi + \Theta_0(\varphi)$ and $\Theta_0(\varphi)$ is a periodic function. Therefore

$$\widehat{\eta} := \mathcal{D}_\omega^{-1}\big(g_2 - [\partial_\psi\theta_0(\varphi)]^T\langle g_2\rangle\big) + \langle\widehat{\eta}\rangle, \quad \langle\widehat{\eta}\rangle \in \mathbb{R}^\nu, \tag{6.31}$$

where the average $\langle\widehat{\eta}\rangle$ will be fixed below. Then we consider the third equation

$$\mathcal{L}_\omega\widehat{w} = g_3 + \partial_x K_{11}(\varphi)\widehat{\eta}, \quad \mathcal{L}_\omega := \omega\cdot\partial_\varphi - \partial_x K_{02}(\varphi). \tag{6.32}$$

- Inversion assumption. *There exists a set* $\Omega_\infty \subset \Omega_o$ *such that for all* $\omega \in \Omega_\infty$, *for every function* $g \in H_{S^\perp}^{s+\mu}(\mathbb{T}^{\nu+1})$ *there exists a solution* $h := \mathcal{L}_\omega^{-1}g \in H_{S^\perp}^s(\mathbb{T}^{\nu+1})$ *of the linear equation* $\mathcal{L}_\omega h = g$ *which satisfies*

$$\|\mathcal{L}_\omega^{-1}g\|_s^{\mathrm{Lip}(\gamma)} \leq C(s)\gamma^{-1}\big(\|g\|_{s+\mu}^{\mathrm{Lip}(\gamma)} + \varepsilon\gamma^{-1}\|\mathfrak{I}_0\|_{s+\mu}^{\mathrm{Lip}(\gamma)}\|g\|_{s_0}^{\mathrm{Lip}(\gamma)}\big) \tag{6.33}$$

for some $\mu := \mu(\tau,\nu) > 0$.

Remark 6.8. The term $\varepsilon\gamma^{-1}\|\mathfrak{I}_0\|_{s+\mu}^{\mathrm{Lip}(\gamma)}$ arises because the remainder R_6 in Section 8.6 contains the term $\varepsilon(\|\Theta_0\|_{s+\mu}^{\mathrm{Lip}(\gamma)} + \|y_\delta\|_{s+\mu}^{\mathrm{Lip}(\gamma)})$ (which is bounded by $\varepsilon\|\mathfrak{I}_0\|_{s+\mu}^{\mathrm{Lip}(\gamma)}$ by (6.9)), see Lemma 8.24.

By the above assumption there exists a solution

$$\widehat{w} := \mathcal{L}_\omega^{-1}[g_3 + \partial_x K_{11}(\varphi)\widehat{\eta}\,] \tag{6.34}$$

of (6.32). Finally, we solve the first equation in (6.29), which, substituting (6.31), (6.34), becomes

$$\mathcal{D}_\omega \widehat{\psi} = g_1 + M_1(\varphi)\langle\widehat{\eta}\rangle + M_2(\varphi)g_2 + M_3(\varphi)g_3 - M_2(\varphi)[\partial_\psi \theta_0]^T \langle g_2\rangle\,, \tag{6.35}$$

where

$$M_1(\varphi) := K_{20}(\varphi) + K_{11}^T(\varphi)\mathcal{L}_\omega^{-1}\partial_x K_{11}(\varphi)\,, \quad M_2(\varphi) := M_1(\varphi)\mathcal{D}_\omega^{-1}\,, \quad M_3(\varphi) := K_{11}^T(\varphi)\mathcal{L}_\omega^{-1}\,. \tag{6.36}$$

In order to solve the equation (6.35) we have to choose $\langle\widehat{\eta}\rangle$ such that the right hand side in (6.35) has zero average. By Lemma 6.6 and (6.4), the φ-averaged matrix $\langle M_1\rangle = -3\varepsilon^{2b} I + O(\varepsilon^{10}\gamma^{-3})$. Therefore, for ε small, $\langle M_1\rangle$ is invertible and $\langle M_1\rangle^{-1} = O(\varepsilon^{-2b}) = O(\gamma^{-1})$ (recall (5.10)). Thus we define

$$\langle\widehat{\eta}\rangle := -\langle M_1\rangle^{-1}[\langle g_1\rangle + \langle M_2 g_2\rangle + \langle M_3 g_3\rangle - \langle M_2[\partial_\psi\theta_0]^T\rangle\langle g_2\rangle]. \tag{6.37}$$

With this choice of $\langle\widehat{\eta}\rangle$ the equation (6.35) has the solution

$$\widehat{\psi} := \mathcal{D}_\omega^{-1}[g_1 + M_1(\varphi)\langle\widehat{\eta}\rangle + M_2(\varphi)g_2 + M_3(\varphi)g_3 - M_2(\varphi)[\partial_\psi\theta_0]^T\langle g_2\rangle]. \tag{6.38}$$

In conclusion, we have constructed a solution $(\widehat{\psi}, \widehat{\eta}, \widehat{w}, \widehat{\zeta})$ of the linear system (6.29).

Proposition 6.9. *Assume (6.4) and (6.33). Then, $\forall \omega \in \Omega_\infty$, $\forall g := (g_1, g_2, g_3)$, the system (6.29) has a solution $\mathbb{D}^{-1} g := (\widehat{\psi}, \widehat{\eta}, \widehat{w}, \widehat{\zeta})$ where $(\widehat{\psi}, \widehat{\eta}, \widehat{w}, \widehat{\zeta})$ are defined in (6.38), (6.31), (6.37), (6.34), (6.30) satisfying*

$$\|\mathbb{D}^{-1} g\|_s^{\mathrm{Lip}(\gamma)} \leq_s \gamma^{-1}\big(\|g\|_{s+\mu}^{\mathrm{Lip}(\gamma)} + \varepsilon\gamma^{-1}\|\mathfrak{I}_0\|_{s+\mu}^{\mathrm{Lip}(\gamma)}\|g\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)}\big). \tag{6.39}$$

Proof. Recalling (6.36), by Lemma 6.6, (6.33), (6.4) we get $\|M_2 h\|_{s_0} + \|M_3 h\|_{s_0} \leq C\|h\|_{s_0+\sigma}$. Then, by (6.37) and $\langle M_1\rangle^{-1} = O(\varepsilon^{-2b}) = O(\gamma^{-1})$, we deduce $|\langle\widehat{\eta}\rangle|^{\mathrm{Lip}(\gamma)} \leq C\gamma^{-1}\|g\|_{s_0+\sigma}^{\mathrm{Lip}(\gamma)}$ and (6.31), (5.21) imply $\|\widehat{\eta}\|_s^{\mathrm{Lip}(\gamma)} \leq_s \gamma^{-1}\big(\|g\|_{s+\sigma}^{\mathrm{Lip}(\gamma)} + \|\mathfrak{I}_0\|_{s+\sigma}\|g\|_{s_0}^{\mathrm{Lip}(\gamma)}\big)$. The bound (6.39) is sharp for $\widehat{w}$ because $\mathcal{L}_\omega^{-1} g_3$ in (6.34) is estimated using (6.33). Finally $\widehat{\psi}$ satisfies (6.39) using (6.38), (6.36), (6.33), (5.21) and Lemma 6.6. $\square$

Finally we prove that the operator

$$\mathbf{T}_0 := (D\widetilde{G}_\delta)(\varphi, 0, 0) \circ \mathbb{D}^{-1} \circ (DG_\delta)(\varphi, 0, 0)^{-1} \tag{6.40}$$

is an approximate right inverse for $d_{i,\zeta}\mathcal{F}(i_0)$ where $\widetilde{G}_\delta(\psi, \eta, w, \zeta) := \big(G_\delta(\psi, \eta, w), \zeta\big)$ is the identity on the ζ-component. We denote the norm $\|(\psi, \eta, w, \zeta)\|_s^{\mathrm{Lip}(\gamma)} := \max\{\|(\psi, \eta, w)\|_s^{\mathrm{Lip}(\gamma)}, |\zeta|^{\mathrm{Lip}(\gamma)}\}$.

Theorem 6.10 *(Approximate inverse). Assume (6.4) and the inversion assumption (6.33). Then there exists $\mu := \mu(\tau, \nu) > 0$ such that, for all $\omega \in \Omega_\infty$, for all $g := (g_1, g_2, g_3)$, the operator $\mathbf{T}_0$ defined in (6.40) satisfies*

$$\|\mathbf{T}_0 g\|_s^{\mathrm{Lip}(\gamma)} \leq_s \gamma^{-1}\big(\|g\|_{s+\mu}^{\mathrm{Lip}(\gamma)} + \varepsilon\gamma^{-1}\|\mathfrak{I}_0\|_{s+\mu}^{\mathrm{Lip}(\gamma)}\|g\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)}\big). \tag{6.41}$$

It is an approximate inverse of $d_{i,\zeta}\mathcal{F}(i_0)$, namely

$$\begin{aligned} &\|(d_{i,\zeta}\mathcal{F}(i_0) \circ \mathbf{T}_0 - I)g\|_s^{\mathrm{Lip}(\gamma)} \\ &\quad \leq_s \varepsilon^{2b-1}\gamma^{-2}\Big(\|\mathcal{F}(i_0, \zeta_0)\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)}\|g\|_{s+\mu}^{\mathrm{Lip}(\gamma)} + \big\{\|\mathcal{F}(i_0, \zeta_0)\|_{s+\mu}^{\mathrm{Lip}(\gamma)} \\ &\qquad + \varepsilon\gamma^{-1}\|\mathcal{F}(i_0, \zeta_0)\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)}\|\mathfrak{I}_0\|_{s+\mu}^{\mathrm{Lip}(\gamma)}\big\}\|g\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)}\Big). \end{aligned} \tag{6.42}$$

Proof. We denote $\| \ \|_s$ instead of $\| \ \|_s^{\mathrm{Lip}(\gamma)}$. The bound (6.41) follows from (6.40), (6.39), (6.28). By (5.6), since $X_{\mathcal{N}}$ does not depend on y, and i_δ differs from i_0 only for the y component, we have

$$d_{i,\zeta}\mathcal{F}(i_0)[\widehat{\imath},\widehat{\zeta}\,]-d_{i,\zeta}\mathcal{F}(i_\delta)[\widehat{\imath},\widehat{\zeta}\,]=d_iX_P(i_\delta)[\widehat{\imath}\,]-d_iX_P(i_0)[\widehat{\imath}\,]$$
$$=\int_0^1\partial_y d_iX_P(\theta_0,y_0+s(y_\delta-y_0),z_0)[y_\delta-y_0,\widehat{\imath}\,]ds=:\mathcal{E}_0[\widehat{\imath},\widehat{\zeta}\,]. \tag{6.43}$$

By (5.18), (6.9), (6.10), (6.4), we estimate

$$\|\mathcal{E}_0[\widehat{\imath},\widehat{\zeta}\,]\|_s\le_s\varepsilon^{2b-1}\gamma^{-1}\Big(\|Z\|_{s_0+\sigma}\|\widehat{\imath}\|_{s+\sigma}+\|Z\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}+\|Z\|_{s_0+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\|\mathfrak{I}_0\|_{s+\sigma}\Big) \tag{6.44}$$

where $Z:=\mathcal{F}(i_0,\zeta_0)$ (recall (6.5)). Note that $\mathcal{E}_0[\widehat{\imath},\widehat{\zeta}]$ is, in fact, independent of $\widehat{\zeta}$. Denote the set of variables $(\psi,\eta,w)=:\mathrm{u}$. Under the transformation G_δ, the nonlinear operator $\mathcal{F}$ in (5.6) transforms into

$$\mathcal{F}(G_\delta(\mathrm{u}(\varphi)),\zeta)=DG_\delta(\mathrm{u}(\varphi))\big(\mathcal{D}_\omega\mathrm{u}(\varphi)-X_K(\mathrm{u}(\varphi),\zeta)\big),\quad K=H_{\varepsilon,\zeta}\circ G_\delta, \tag{6.45}$$

see (6.21). Differentiating (6.45) at the trivial torus $\mathrm{u}_\delta(\varphi)=G_\delta^{-1}(i_\delta)(\varphi)=(\varphi,0,0)$, at $\zeta=\zeta_0$, in the directions $(\widehat{\mathrm{u}},\widehat{\zeta})=(DG_\delta(\mathrm{u}_\delta)^{-1}[\widehat{\imath}\,],\widehat{\zeta})=D\widetilde{G}_\delta(\mathrm{u}_\delta)^{-1}[\widehat{\imath},\widehat{\zeta}\,]$, we get

$$d_{i,\zeta}\mathcal{F}(i_\delta)[\widehat{\imath},\widehat{\zeta}\,]=DG_\delta(\mathrm{u}_\delta)\big(\mathcal{D}_\omega\widehat{\mathrm{u}}-d_{\mathrm{u},\zeta}X_K(\mathrm{u}_\delta,\zeta_0)[\widehat{\mathrm{u}},\widehat{\zeta}\,]\big)+\mathcal{E}_1[\widehat{\imath},\widehat{\zeta}\,], \tag{6.46}$$
$$\mathcal{E}_1[\widehat{\imath},\widehat{\zeta}\,]:=D^2G_\delta(\mathrm{u}_\delta)\big[DG_\delta(\mathrm{u}_\delta)^{-1}\mathcal{F}(i_\delta,\zeta_0),\ DG_\delta(\mathrm{u}_\delta)^{-1}[\widehat{\imath}\,]\big], \tag{6.47}$$

where $d_{\mathrm{u},\zeta}X_K(\mathrm{u}_\delta,\zeta_0)$ is expanded in (6.27). In fact, $\mathcal{E}_1$ is independent of $\widehat{\zeta}$. We split

$$\mathcal{D}_\omega\widehat{\mathrm{u}}-d_{\mathrm{u},\zeta}X_K(\mathrm{u}_\delta,\zeta_0)[\widehat{\mathrm{u}},\widehat{\zeta}\,]=\mathbb{D}[\widehat{\mathrm{u}},\widehat{\zeta}\,]+R_Z[\widehat{\mathrm{u}},\widehat{\zeta}\,],$$

where $\mathbb{D}[\widehat{\mathrm{u}},\widehat{\zeta}]$ is defined in (6.29) and

$$R_Z[\widehat{\psi},\widehat{\eta},\widehat{w},\widehat{\zeta}\,]:=\begin{pmatrix}-\partial_\psi K_{10}(\varphi)[\widehat{\psi}]\\ \partial_\psi[\partial_\psi\theta_0(\varphi)]^T[\widehat{\psi},\zeta_0]+\partial_{\psi\psi}K_{00}(\varphi)[\widehat{\psi}]+[\partial_\psi K_{10}(\varphi)]^T\widehat{\eta}+[\partial_\psi K_{01}(\varphi)]^T\widehat{w}\\ -\partial_x\{\partial_\psi K_{01}(\varphi)[\widehat{\psi}]\}\end{pmatrix} \tag{6.48}$$

(R_Z is independent of $\widehat{\zeta}$). By (6.43) and (6.46),

$$d_{i,\zeta}\mathcal{F}(i_0)=DG_\delta(\mathrm{u}_\delta)\circ\mathbb{D}\circ D\widetilde{G}_\delta(\mathrm{u}_\delta)^{-1}+\mathcal{E}_0+\mathcal{E}_1+\mathcal{E}_2,\quad \mathcal{E}_2:=DG_\delta(\mathrm{u}_\delta)\circ R_Z\circ D\widetilde{G}_\delta(\mathrm{u}_\delta)^{-1}. \tag{6.49}$$

By Lemmata 6.4, 6.7, 6.1, and (6.11), (6.4), the terms $\mathcal{E}_1$, $\mathcal{E}_2$ (see (6.47), (6.49), (6.48)) satisfy the same bound (6.44) as $\mathcal{E}_0$ (in fact even better). Thus the sum $\mathcal{E}:=\mathcal{E}_0+\mathcal{E}_1+\mathcal{E}_2$ satisfies (6.44). Applying $\mathbf{T}_0$ defined in (6.40) to the right in (6.49), since $\mathbb{D}\circ\mathbb{D}^{-1}=I$ (see Proposition 6.9), we get $d_{i,\zeta}\mathcal{F}(i_0)\circ\mathbf{T}_0-I=\mathcal{E}\circ\mathbf{T}_0$. Then (6.42) follows from (6.41) and the bound (6.44) for $\mathcal{E}$. □

7. The linearized operator in the normal directions

The goal of this section is to write an explicit expression of the linearized operator $\mathcal{L}_\omega$ defined in (6.32), see Proposition 7.6. To this aim, we compute $\frac{1}{2}(K_{02}(\psi)w,w)_{L^2(\mathbb{T})}$, $w\in H_S^\perp$, which collects all the components of $(H_\varepsilon\circ G_\delta)(\psi,0,w)$ that are quadratic in w, see (6.20).

We first prove some preliminary lemmata.

Lemma 7.1. *Let H be a Hamiltonian of class $C^2(H_0^1(\mathbb{T}_x),\mathbb{R})$ and consider a map $\Phi(u):=u+\Psi(u)$ satisfying $\Psi(u)=\Pi_E\Psi(\Pi_Eu)$, for all u, where E is a finite dimensional subspace as in (3.3). Then*

$$\partial_u\big[\nabla(H\circ\Phi)\big](u)[h]=(\partial_u\nabla H)(\Phi(u))[h]+\mathcal{R}(u)[h], \tag{7.1}$$

where $\mathcal{R}(u)$ has the "finite dimensional" form

$$\mathcal{R}(u)[h]=\sum_{|j|\le C}\big(h,g_j(u)\big)_{L^2(\mathbb{T})}\chi_j(u) \tag{7.2}$$

with $\chi_j(u) = e^{\mathrm{i}jx}$ *or* $g_j(u) = e^{\mathrm{i}jx}$. *The remainder* $\mathcal{R}(u) = \mathcal{R}_0(u) + \mathcal{R}_1(u) + \mathcal{R}_2(u)$ *with*

$$\mathcal{R}_0(u) := (\partial_u \nabla H)(\Phi(u))\partial_u \Psi(u), \qquad \mathcal{R}_1(u) := [\partial_u\{\Psi'(u)^T\}][\cdot, \nabla H(\Phi(u))],$$
$$\mathcal{R}_2(u) := [\partial_u \Psi(u)]^T (\partial_u \nabla H)(\Phi(u))\partial_u \Phi(u). \tag{7.3}$$

Proof. By a direct calculation,

$$\nabla(H \circ \Phi)(u) = [\Phi'(u)]^T \nabla H(\Phi(u)) = \nabla H(\Phi(u)) + [\Psi'(u)]^T \nabla H(\Phi(u)) \tag{7.4}$$

where $\Phi'(u) := (\partial_u \Phi)(u)$ and $[\ ]^T$ denotes the transpose with respect to the L^2 scalar product. Differentiating (7.4), we get (7.1) and (7.3).

Let us show that each $\mathcal{R}_m$ has the form (7.2). We have

$$\Psi'(u) = \Pi_E \Psi'(\Pi_E u)\Pi_E\ , \quad [\Psi'(u)]^T = \Pi_E[\Psi'(\Pi_E u)]^T \Pi_E\ . \tag{7.5}$$

Hence, setting $A := (\partial_u \nabla H)(\Phi(u))\Pi_E \Psi'(\Pi_E u)$, we get

$$\mathcal{R}_0(u)[h] = A[\Pi_E h] = \sum_{|j|\le C} h_j A(e^{\mathrm{i}jx}) = \sum_{|j|\le C} (h, g_j)_{L^2(\mathbb{T})} \chi_j$$

with $g_j := e^{\mathrm{i}jx}$, $\chi_j := A(e^{\mathrm{i}jx})$. Similarly, using (7.5), and setting $A := [\Psi'(\Pi_E u)]^T \Pi_E (\partial_u \nabla H)(\Phi(u))\Phi'(u)$, we get

$$\mathcal{R}_2(u)[h] = \Pi_E[Ah] = \sum_{|j|\le C} (Ah, e^{\mathrm{i}jx})_{L^2(\mathbb{T})} e^{\mathrm{i}jx} = \sum_{|j|\le C} (h, A^T e^{\mathrm{i}jx})_{L^2(\mathbb{T})} e^{\mathrm{i}jx}\ ,$$

which has the form (7.2) with $g_j := A^T(e^{\mathrm{i}jx})$ and $\chi_j := e^{\mathrm{i}jx}$. Differentiating the second equality in (7.5), we see that

$$\mathcal{R}_1(u)[h] = \Pi_E[Ah]\,, \quad Ah := \partial_u\{\Psi'(\Pi_E u)^T\}[\Pi_E h, \Pi_E(\nabla H)(\Phi(u))]\,,$$

which has the same form of $\mathcal{R}_2$ and so (7.2). □

Lemma 7.2. *Let* $H(u) := \int_{\mathbb{T}} f(u)X(u)dx$ *where* $X(u) = \Pi_E X(\Pi_E u)$ *and* $f(u)(x) := f(u(x))$ *is the composition operator for a function of class* C^2. *Then*

$$(\partial_u \nabla H)(u)[h] = f''(u)X(u)h + \mathcal{R}(u)[h] \tag{7.6}$$

where $\mathcal{R}(u)$ *has the form* (7.2) *with* $\chi_j(u) = e^{\mathrm{i}jx}$ *or* $g_j(u) = e^{\mathrm{i}jx}$.

Proof. A direct calculation proves that $\nabla H(u) = f'(u)X(u) + X'(u)^T[f(u)]$, and (7.6) follows with $\mathcal{R}(u)[h] = f'(u)X'(u)[h] + \partial_u\{X'(u)^T\}[h, f(u)] + X'(u)^T[f'(u)h]$, which has the form (7.2). □

We conclude this section with a technical lemma used from the end of Section 8.3 about the decay norms of "finite dimensional operators". Note that operators of the form (7.7) (that will appear in Section 8.1) reduce to those in (7.2) when the functions $g_j(\tau)$, $\chi_j(\tau)$ are independent of τ.

Lemma 7.3. *Let* $\mathcal{R}$ *be an operator of the form*

$$\mathcal{R}h = \sum_{|j|\le C} \int_0^1 \big(h\,,\, g_j(\tau)\big)_{L^2(\mathbb{T})} \chi_j(\tau)\, d\tau\,, \tag{7.7}$$

where the functions $g_j(\tau)$, $\chi_j(\tau) \in H^s$, $\tau \in [0, 1]$ *depend in a Lipschitz way on the parameter* ω. *Then its matrix* s*-decay norm (see* (2.16)–(2.17)*) satisfies*

$$|\mathcal{R}|_s^{\mathrm{Lip}(\gamma)} \le_s \sum_{|j|\le C} \sup_{\tau\in[0,1]}\big\{\|\chi_j(\tau)\|_s^{\mathrm{Lip}(\gamma)} \|g_j(\tau)\|_{s_0}^{\mathrm{Lip}(\gamma)} + \|\chi_j(\tau)\|_{s_0}^{\mathrm{Lip}(\gamma)} \|g_j(\tau)\|_s^{\mathrm{Lip}(\gamma)}\big\}\,.$$

Proof. For each $\tau \in [0, 1]$, the operator $h \mapsto (h, g_j(\tau))\chi_j(\tau)$ is the composition $\chi_j(\tau) \circ \Pi_0 \circ g_j(\tau)$ of the multiplication operators for $g_j(\tau)$, $\chi_j(\tau)$ and $h \mapsto \Pi_0 h := \int_{\mathbb{T}} h dx$. Hence the lemma follows by the interpolation estimate (2.20) and (2.18). □

7.1. Composition with the map G_δ

In the sequel we shall use that $\mathfrak{I}_\delta := \mathfrak{I}_\delta(\varphi;\omega) := i_\delta(\varphi;\omega) - (\varphi, 0, 0)$ satisfies, by (6.9) and (6.4),

$$\|\mathfrak{I}_\delta\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)} \leq C\varepsilon^{6-2b}\gamma^{-1}\,. \tag{7.8}$$

We now study the Hamiltonian $K := H_\varepsilon \circ G_\delta = \varepsilon^{-2b}\mathcal{H} \circ A_\varepsilon \circ G_\delta$ defined in (6.20), (4.6).

Recalling (4.7) and (6.19) the map $A_\varepsilon \circ G_\delta$ has the form

$$A_\varepsilon \circ G_\delta(\psi,\eta,w) = \varepsilon \sum_{j\in S} \sqrt{\xi_j + \varepsilon^{2(b-1)}|j|[y_\delta(\psi) + L_1(\psi)\eta + L_2(\psi)w]_j}\, e^{\mathrm{i}[\theta_0(\psi)]_j} e^{\mathrm{i}jx} + \varepsilon^b(z_0(\psi) + w) \tag{7.9}$$

where

$$L_1(\psi) := [\partial_\psi \theta_0(\psi)]^{-T}\,, \quad L_2(\psi) := \big[(\partial_\theta \tilde{z}_0)(\theta_0(\psi))\big]^T \partial_x^{-1}\,. \tag{7.10}$$

By Taylor's formula, we develop (7.9) in w at $\eta = 0$, $w = 0$, and we get $A_\varepsilon \circ G_\delta(\psi, 0, w) = T_\delta(\psi) + T_1(\psi)w + T_2(\psi)[w,w] + T_{\geq 3}(\psi, w)$, where

$$\begin{aligned} &T_\delta(\psi) := (A_\varepsilon \circ G_\delta)(\psi, 0, 0) = \varepsilon v_\delta(\psi) + \varepsilon^b z_0(\psi)\,, \\ &v_\delta(\psi) := \sum_{j\in S} \sqrt{\xi_j + \varepsilon^{2(b-1)}|j|[y_\delta(\psi)]_j}\, e^{\mathrm{i}[\theta_0(\psi)]_j} e^{\mathrm{i}jx} \end{aligned} \tag{7.11}$$

is the approximate isotropic torus in phase space (it corresponds to i_δ in Lemma 6.3),

$$T_1(\psi)w = \varepsilon \sum_{j\in S} \frac{\varepsilon^{2(b-1)}|j|[L_2(\psi)w]_j\, e^{\mathrm{i}[\theta_0(\psi)]_j}}{2\sqrt{\xi_j + \varepsilon^{2(b-1)}|j|[y_\delta(\psi)]_j}}\, e^{\mathrm{i}jx} + \varepsilon^b w =: \varepsilon^{2b-1}U_1(\psi)w + \varepsilon^b w \tag{7.12}$$

$$T_2(\psi)[w,w] = -\varepsilon \sum_{j\in S} \frac{\varepsilon^{4(b-1)} j^2 [L_2(\psi)w]_j^2\, e^{\mathrm{i}[\theta_0(\psi)]_j}}{8\{\xi_j + \varepsilon^{2(b-1)}|j|[y_\delta(\psi)]_j\}^{3/2}}\, e^{\mathrm{i}jx} =: \varepsilon^{4b-3}U_2(\psi)[w,w] \tag{7.13}$$

and $T_{\geq 3}(\psi, w)$ collects all the terms of order at least cubic in w. In the notation of (4.7), the function $v_\delta(\psi)$ in (7.11) is $v_\delta(\psi) = v_\varepsilon(\theta_0(\psi), y_\delta(\psi))$. The terms $U_1, U_2 = O(1)$ in ε. Moreover, using that $L_2(\psi)$ in (7.10) vanishes as $z_0 = 0$, they satisfy

$$\|U_1 w\|_s \leq \|\mathfrak{I}_\delta\|_s \|w\|_{s_0} + \|\mathfrak{I}_\delta\|_{s_0}\|w\|_s\,, \quad \|U_2[w,w]\|_s \leq \|\mathfrak{I}_\delta\|_s \|\mathfrak{I}_\delta\|_{s_0}\|w\|_{s_0}^2 + \|\mathfrak{I}_\delta\|_{s_0}^2\|w\|_{s_0}\|w\|_s \tag{7.14}$$

and also in the $\|\ \|_s^{\mathrm{Lip}(\gamma)}$-norm.

By Taylor's formula $\mathcal{H}(u+h) = \mathcal{H}(u) + ((\nabla\mathcal{H})(u), h)_{L^2(\mathbb{T})} + \frac{1}{2}((\partial_u\nabla\mathcal{H})(u)[h], h)_{L^2(\mathbb{T})} + O(h^3)$. Specifying at $u = T_\delta(\psi)$ and $h = T_1(\psi)w + T_2(\psi)[w,w] + T_{\geq 3}(\psi, w)$, we obtain that the sum of all the components of $K = \varepsilon^{-2b}(\mathcal{H} \circ A_\varepsilon \circ G_\delta)(\psi, 0, w)$ that are quadratic in w is

$$\frac{1}{2}(K_{02}w, w)_{L^2(\mathbb{T})} = \varepsilon^{-2b}((\nabla\mathcal{H})(T_\delta), T_2[w,w])_{L^2(\mathbb{T})} + \varepsilon^{-2b}\frac{1}{2}((\partial_u\nabla\mathcal{H})(T_\delta)[T_1 w], T_1 w)_{L^2(\mathbb{T})}\,.$$

Inserting the expressions (7.12), (7.13) we get

$$\begin{aligned} K_{02}(\psi)w = {} & (\partial_u\nabla\mathcal{H})(T_\delta)[w] + 2\varepsilon^{b-1}(\partial_u\nabla\mathcal{H})(T_\delta)[U_1 w] + \varepsilon^{2(b-1)}U_1^T(\partial_u\nabla\mathcal{H})(T_\delta)[U_1 w] \\ & + 2\varepsilon^{2b-3}U_2[w,\cdot]^T(\nabla\mathcal{H})(T_\delta). \end{aligned} \tag{7.15}$$

Lemma 7.4.

$$(K_{02}(\psi)w, w)_{L^2(\mathbb{T})} = ((\partial_u\nabla\mathcal{H})(T_\delta)[w], w)_{L^2(\mathbb{T})} + (R(\psi)w, w)_{L^2(\mathbb{T})} \tag{7.16}$$

where $R(\psi)w$ has the "finite dimensional" form

$$R(\psi)w = \sum_{|j|\leq C} \big(w, g_j(\psi)\big)_{L^2(\mathbb{T})}\chi_j(\psi) \tag{7.17}$$

where, for some $\sigma := \sigma(\nu,\tau) > 0$,

$$\|g_j\|_s^{\mathrm{Lip}(\gamma)}\|\chi_j\|_{s_0}^{\mathrm{Lip}(\gamma)} + \|g_j\|_{s_0}^{\mathrm{Lip}(\gamma)}\|\chi_j\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^{b+1}\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)} \tag{7.18}$$

$$\|\partial_i g_j[\widehat{\imath}]\|_s\|\chi_j\|_{s_0} + \|\partial_i g_j[\widehat{\imath}]\|_{s_0}\|\chi_j\|_s + \|g_j\|_{s_0}\|\partial_i\chi_j[\widehat{\imath}]\|_s + \|g_j\|_s\|\partial_i\chi_j[\widehat{\imath}]\|_{s_0}$$
$$\le_s \varepsilon^{b+1}\|\widehat{\imath}\|_{s+\sigma} + \varepsilon^{2b-1}\|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\,, \tag{7.19}$$

and, as usual, $i = (\theta, y, z)$ (see (5.1)), $\widehat{\imath} = (\widehat{\theta}, \widehat{y}, \widehat{z})$.

Proof. Since $U_1 = \Pi_S U_1$ and $U_2 = \Pi_S U_2$, the last three terms in (7.15) have all the form (7.17) (argue as in Lemma 7.1). We now prove that they are also small in size.

The contributions in (7.15) from H_2 are better analyzed by the expression

$$\varepsilon^{-2b} H_2 \circ A_\varepsilon \circ G_\delta(\psi,\eta,w) = const + \sum_{j\in S^+} j^3\big[y_\delta(\psi) + L_1(\psi)\eta + L_2(\psi)w\big]_j + \frac{1}{2}\int_{\mathbb{T}}(z_0(\psi)+w)_x^2\,dx$$

which follows by (4.8), (6.19), (7.10). Hence the only contribution to $(K_{02}w, w)$ is $\int_{\mathbb{T}} w_x^2\,dx$. Now we consider the cubic term $\mathcal{H}_3$ in (3.6). A direct calculation shows that for $u = v + z$, $\nabla\mathcal{H}_3(u) = 3z^2 + 6\Pi_S^\perp(vz)$, and $\partial_u\nabla\mathcal{H}_3(u)[U_1 w] = 6\Pi_S^\perp(zU_1 w)$ (since $U_1 w \in H_S$). Therefore

$$\nabla\mathcal{H}_3(T_\delta) = 3\varepsilon^{2b}z_0^2 + 6\varepsilon^{b+1}\Pi_S^\perp(v_\delta z_0)\,, \quad \partial_u\nabla\mathcal{H}_3(T_\delta)[U_1 w] = 6\varepsilon^b\Pi_S^\perp(z_0\,U_1 w)\,. \tag{7.20}$$

By (7.20) one has $((\partial_u\nabla\mathcal{H}_3)(T_\delta)[U_1 w], U_1 w)_{L^2(\mathbb{T})} = 0$, and since also $U_2 = \Pi_S U_2$,

$$\varepsilon^{b-1}\partial_u\nabla\mathcal{H}_3(T_\delta)[U_1 w] + \varepsilon^{2b-3}U_2[w,\cdot]^T\nabla\mathcal{H}_3(T_\delta) = 6\varepsilon^{2b-1}\Pi_S^\perp(z_0 U_1 w) + 3\varepsilon^{4b-3}U_2[w,\cdot]^T z_0^2\,. \tag{7.21}$$

These terms have the form (7.17) and, using (7.14), (6.4), they satisfy (7.18).

Finally we consider all the terms which arise from $\mathcal{H}_{\ge 4} = O(u^4)$. The operators $\varepsilon^{b-1}\partial_u\nabla\mathcal{H}_{\ge 4}(T_\delta)U_1$, $\varepsilon^{2(b-1)}U_1^T(\partial_u\nabla\mathcal{H}_{\ge 4})(T_\delta)U_1$, $\varepsilon^{2b-3}U_2^T\nabla\mathcal{H}_{\ge 4}(T_\delta)$ have the form (7.17) and, using $\|T_\delta\|_s^{\mathrm{Lip}(\gamma)} \le \varepsilon(1 + \|\mathfrak{I}_\delta\|_s^{\mathrm{Lip}(\gamma)})$, (7.14), (6.4), the bound (7.18) holds. Notice that the biggest term is $\varepsilon^{b-1}\partial_u\nabla\mathcal{H}_{\ge 4}(T_\delta)U_1$.

By (6.12) and using explicit formulae (7.10)–(7.13) we get estimate (7.19). □

The conclusion of this section is that, after the composition with the action-angle variables, the rescaling (4.5), and the transformation G_δ, the linearized operator to analyze is $H_S^\perp \ni w \mapsto (\partial_u\nabla\mathcal{H})(T_\delta)[w]$, up to finite dimensional operators which have the form (7.17) and size (7.18).

7.2. *The linearized operator in the normal directions*

In view of (7.16) we now compute $((\partial_u\nabla\mathcal{H})(T_\delta)[w], w)_{L^2(\mathbb{T})}$, $w \in H_S^\perp$, where $\mathcal{H} = H \circ \Phi_B$ and Φ_B is the Birkhoff map of Proposition 3.1. It is convenient to estimate separately the terms in

$$\mathcal{H} = H \circ \Phi_B = (H_2 + H_3)\circ\Phi_B + H_{\ge 5}\circ\Phi_B \tag{7.22}$$

where H_2, H_3, $H_{\ge 5}$ are defined in (3.1).

We first consider $H_{\ge 5}\circ\Phi_B$. By (3.1) we get $\nabla H_{\ge 5}(u) = \pi_0[(\partial_u f)(x,u,u_x)] - \partial_x\{(\partial_{u_x} f)(x,u,u_x)\}$, see (2.2). Since the Birkhoff transformation Φ_B has the form (3.4), Lemma 7.1 (at $u = T_\delta$, see (7.11)) implies that

$$\partial_u\nabla(H_{\ge 5}\circ\Phi_B)(T_\delta)[h] = (\partial_u\nabla H_{\ge 5})(\Phi_B(T_\delta))[h] + \mathcal{R}_{H_{\ge 5}}(T_\delta)[h]$$
$$= \partial_x(r_1(T_\delta)\partial_x h) + r_0(T_\delta)h + \mathcal{R}_{H_{\ge 5}}(T_\delta)[h] \tag{7.23}$$

where the multiplicative functions $r_0(T_\delta)$, $r_1(T_\delta)$ are

$$r_0(T_\delta) := \sigma_0(\Phi_B(T_\delta)), \qquad \sigma_0(u) := (\partial_{uu} f)(x,u,u_x) - \partial_x\{(\partial_{uu_x} f)(x,u,u_x)\}, \tag{7.24}$$

$$r_1(T_\delta) := \sigma_1(\Phi_B(T_\delta)), \qquad \sigma_1(u) := -(\partial_{u_x u_x} f)(x,u,u_x), \tag{7.25}$$

the remainder $\mathcal{R}_{H_{\ge 5}}(u)$ has the form (7.2) with $\chi_j = e^{\mathrm{i}jx}$ or $g_j = e^{\mathrm{i}jx}$ and, using (7.3), it satisfies, for some $\sigma := \sigma(\nu,\tau) > 0$,

$$\|g_j\|_s^{\mathrm{Lip}(\gamma)}\|\chi_j\|_{s_0}^{\mathrm{Lip}(\gamma)} + \|g_j\|_{s_0}^{\mathrm{Lip}(\gamma)}\|\chi_j\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^4(1+\|\mathfrak{I}_\delta\|_{s+2}^{\mathrm{Lip}(\gamma)})$$
$$\|\partial_i g_j[\widehat{\imath}]\|_s\|\chi_j\|_{s_0} + \|\partial_i g_j[\widehat{\imath}]\|_{s_0}\|\chi_j\|_s + \|g_j\|_{s_0}\|\partial_i\chi_j[\widehat{\imath}]\|_s + \|g_j\|_s\|\partial_i\chi_j[\widehat{\imath}]\|_{s_0}$$
$$\le_s \varepsilon^4(\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+2}\|\widehat{\imath}\|_{s_0+2}).$$

Now we consider the contributions from $(H_2+H_3)\circ\Phi_B$. By Lemma 7.1 and the expressions of H_2, H_3 in (3.1) we deduce that

$$\partial_u\nabla(H_2\circ\Phi_B)(T_\delta)[h] = -\partial_{xx}h + \mathcal{R}_{H_2}(T_\delta)[h], \quad \partial_u\nabla(H_3\circ\Phi_B)(T_\delta)[h] = 6\Phi_B(T_\delta)h + \mathcal{R}_{H_3}(T_\delta)[h],$$

where $\Phi_B(T_\delta)$ is a function with zero space average, because $\Phi_B : H_0^1(\mathbb{T}_x)\to H_0^1(\mathbb{T}_x)$ (Proposition 3.1) and $\mathcal{R}_{H_2}(u)$, $\mathcal{R}_{H_3}(u)$ have the form (7.2). By (7.3), the size $(\mathcal{R}_{H_2}+\mathcal{R}_{H_3})(T_\delta) = O(\varepsilon)$. We expand

$$(\mathcal{R}_{H_2}+\mathcal{R}_{H_3})(T_\delta) = \varepsilon\mathcal{R}_1 + \varepsilon^2\mathcal{R}_2 + \tilde{\mathcal{R}}_{>2},$$

where $\tilde{\mathcal{R}}_{>2}$ has size $o(\varepsilon^2)$, and we get, $\forall h\in H_S^\perp$,

$$\Pi_S^\perp\partial_u\nabla((H_2+H_3)\circ\Phi_B)(T_\delta)[h] = -\partial_{xx}h + \Pi_S^\perp(6\Phi_B(T_\delta)h) + \Pi_S^\perp(\varepsilon\mathcal{R}_1 + \varepsilon^2\mathcal{R}_2 + \tilde{\mathcal{R}}_{>2})[h]. \tag{7.26}$$

We also develop the function $\Phi_B(T_\delta)$ is powers of ε. Expand $\Phi_B(u) = u + \Psi_2(u) + \Psi_{\ge 3}(u)$, where $\Psi_2(u)$ is quadratic, $\Psi_{\ge 3}(u) = O(u^3)$, and both map $H_0^1(\mathbb{T}_x)\to H_0^1(\mathbb{T}_x)$. At $u = T_\delta = \varepsilon v_\delta + \varepsilon^b z_0$ we get

$$\Phi_B(T_\delta) = T_\delta + \Psi_2(T_\delta) + \Psi_{\ge 3}(T_\delta) = \varepsilon v_\delta + \varepsilon^2\Psi_2(v_\delta) + \tilde{q} \tag{7.27}$$

where $\tilde{q} := \varepsilon^b z_0 + \Psi_2(T_\delta) - \varepsilon^2\Psi_2(v_\delta) + \Psi_{\ge 3}(T_\delta)$ has zero space average and it satisfies

$$\|\tilde{q}\|_s^{\mathrm{Lip}(\gamma)} \le_s \varepsilon^3 + \varepsilon^b\|\mathfrak{I}_\delta\|_s^{\mathrm{Lip}(\gamma)}, \quad \|\partial_i\tilde{q}[\widehat{\imath}]\|_s \le_s \varepsilon^b\big(\|\widehat{\imath}\|_s + \|\mathfrak{I}_\delta\|_s\|\widehat{\imath}\|_{s_0}\big).$$

In particular, its low norm $\|\tilde{q}\|_{s_0}^{\mathrm{Lip}(\gamma)} \le_{s_0} \varepsilon^{6-b}\gamma^{-1} = o(\varepsilon^2)$.

We need an exact expression of the terms of order ε and ε^2 in (7.26). We compare the Hamiltonian (3.5) with (7.22), noting that $(H_{\ge 5}\circ\Phi_B)(u) = O(u^5)$ because f satisfies (1.5) and $\Phi_B(u) = O(u)$. Therefore

$$(H_2+H_3)\circ\Phi_B = H_2 + \mathcal{H}_3 + \mathcal{H}_4 + O(u^5),$$

and the homogeneous terms of $(H_2+H_3)\circ\Phi_B$ of degree 2, 3, 4 in u are H_2, $\mathcal{H}_3$, $\mathcal{H}_4$ respectively. As a consequence, the terms of order ε and ε^2 in (7.26) (both in the function $\Phi_B(T_\delta)$ and in the remainders $\mathcal{R}_1$, $\mathcal{R}_2$) come only from $H_2+\mathcal{H}_3+\mathcal{H}_4$. Actually they come from H_2, $\mathcal{H}_3$ and $\mathcal{H}_{4,2}$ (see (3.6), (3.7)) because, at $u = T_\delta = \varepsilon v_\delta + \varepsilon^b z_0$, for all $h\in H_S^\perp$,

$$\Pi_S^\perp(\partial_u\nabla\mathcal{H}_4)(T_\delta)[h] = \Pi_S^\perp(\partial_u\nabla\mathcal{H}_{4,2})(T_\delta)[h] + o(\varepsilon^2).$$

A direct calculation based on the expressions (3.6), (3.7) shows that, for all $h\in H_S^\perp$,

$$\begin{aligned}\Pi_S^\perp(\partial_u\nabla(H_2+\mathcal{H}_3+\mathcal{H}_4))(T_\delta)[h] = -\partial_{xx}h + 6\varepsilon\Pi_S^\perp(v_\delta h) + 6\varepsilon^b\Pi_S^\perp(z_0h) + \varepsilon^2\Pi_S^\perp\big\{6\pi_0[(\partial_x^{-1}v_\delta)^2]h \\ + 6v_\delta\Pi_S[(\partial_x^{-1}v_\delta)(\partial_x^{-1}h)] - 6\partial_x^{-1}\{(\partial_x^{-1}v_\delta)\Pi_S[v_\delta h]\}\big\} + o(\varepsilon^2).\end{aligned} \tag{7.28}$$

Thus, comparing the terms of order ε, ε^2 in (7.26) (using (7.27)) with those in (7.28) we deduce that the operators $\mathcal{R}_1$, $\mathcal{R}_2$ and the function $\Psi_2(v_\delta)$ are

$$\mathcal{R}_1 = 0, \quad \mathcal{R}_2[h] = 6v_\delta\Pi_S\big[(\partial_x^{-1}v_\delta)(\partial_x^{-1}h)\big] - 6\partial_x^{-1}\{(\partial_x^{-1}v_\delta)\Pi_S[v_\delta h]\}, \quad \Psi_2(v_\delta) = \pi_0[(\partial_x^{-1}v_\delta)^2]. \tag{7.29}$$

In conclusion, by (7.22), (7.26), (7.23), (7.27), (7.29), we get, for all $h\in H_{S^\perp}$,

$$\begin{aligned}\Pi_S^\perp\partial_u\nabla\mathcal{H}(T_\delta)[h] = -\partial_{xx}h + \Pi_S^\perp\big[\big(\varepsilon 6v_\delta + \varepsilon^2 6\pi_0[(\partial_x^{-1}v_\delta)^2] + q_{>2} + p_{\ge 4}\big)h\big] \\ + \Pi_S^\perp\partial_x(r_1(T_\delta)\partial_x h) + \varepsilon^2\Pi_S^\perp\mathcal{R}_2[h] + \Pi_S^\perp\mathcal{R}_{>2}[h]\end{aligned} \tag{7.30}$$

where r_1 is defined in (7.24), $\mathcal{R}_2$ in (7.29), the remainder $\mathcal{R}_{>2} := \tilde{\mathcal{R}}_{>2} + \mathcal{R}_{H_{\ge 5}}(T_\delta)$ and the functions (using also (7.24), (7.25), (1.5)),

$$q_{>2} := 6\tilde{q} + \varepsilon^3\big((\partial_{uu}f_5)(v_\delta,(v_\delta)_x) - \partial_x\{(\partial_{uu_x}f_5)(v_\delta,(v_\delta)_x)\}\big) \tag{7.31}$$

$$p_{\ge 4} := r_0(T_\delta) - \varepsilon^3\big[(\partial_{uu}f_5)(v_\delta,(v_\delta)_x) - \partial_x\{(\partial_{uu_x}f_5)(v_\delta,(v_\delta)_x)\}\big]. \tag{7.32}$$

Lemma 7.5. $\int_{\mathbb{T}} q_{>2} dx = 0$.

Proof. We already observed that $\tilde{q}$ has zero x-average as well as the derivative $\partial_x\{(\partial_{uu_x} f_5)(v, v_x)\}$. Finally

$$(\partial_{uu} f_5)(v, v_x) = \sum_{j_1, j_2, j_3 \in S} c_{j_1 j_2 j_3} v_{j_1} v_{j_2} v_{j_3} e^{\mathrm{i}(j_1+j_2+j_3)x}, \quad v := \sum_{j \in S} v_j e^{\mathrm{i} j x} \tag{7.33}$$

for some coefficient $c_{j_1 j_2 j_3}$, and therefore it has zero average by hypothesis $(\mathbb{S}1)$. $\square$

By Lemma 7.4 and the results of this section (in particular (7.30)) we deduce:

Proposition 7.6. *Assume (7.8). Then the Hamiltonian operator $\mathcal{L}_\omega$ has the form, $\forall h \in H^s_{S^\perp}(\mathbb{T}^{\nu+1})$,*

$$\mathcal{L}_\omega h := \omega \cdot \partial_\varphi h - \partial_x K_{02} h = \Pi_S^\perp \big(\omega \cdot \partial_\varphi h + \partial_{xx}(a_1 \partial_x h) + \partial_x(a_0 h) - \varepsilon^2 \partial_x \mathcal{R}_2 h - \partial_x \mathcal{R}_* h\big) \tag{7.34}$$

where $\mathcal{R}_2$ is defined in (7.29), $\mathcal{R}_* := \mathcal{R}_{>2} + R(\psi)$ (with $R(\psi)$ defined in Lemma 7.4), the functions

$$a_1 := 1 - r_1(T_\delta), \quad a_0 := -(\varepsilon p_1 + \varepsilon^2 p_2 + q_{>2} + p_{\geq 4}), \quad p_1 := 6 v_\delta, \quad p_2 := 6\pi_0[(\partial_x^{-1} v_\delta)^2], \tag{7.35}$$

the function $q_{>2}$ is defined in (7.31) and satisfies $\int_{\mathbb{T}} q_{>2} dx = 0$, the function $p_{\geq 4}$ is defined in (7.32), r_1 in (7.25), T_δ and v_δ in (7.11). For $p_k = p_1, p_2$,

$$\|p_k\|_s^{\mathrm{Lip}(\gamma)} \leq_s 1 + \|\mathfrak{I}_\delta\|_s^{\mathrm{Lip}(\gamma)}, \qquad \|\partial_i p_k[\hat{\imath}]\|_s \leq_s \|\hat{\imath}\|_{s+1} + \|\mathfrak{I}_\delta\|_{s+1}\|\hat{\imath}\|_{s_0+1}, \tag{7.36}$$

$$\|q_{>2}\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^3 + \varepsilon^b \|\mathfrak{I}_\delta\|_s^{\mathrm{Lip}(\gamma)}, \qquad \|\partial_i q_{>2}[\hat{\imath}]\|_s \leq_s \varepsilon^b\big(\|\hat{\imath}\|_{s+1} + \|\mathfrak{I}_\delta\|_{s+1}\|\hat{\imath}\|_{s_0+1}\big), \tag{7.37}$$

$$\|a_1 - 1\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^3\big(1 + \|\mathfrak{I}_\delta\|_{s+1}^{\mathrm{Lip}(\gamma)}\big), \qquad \|\partial_i a_1[\hat{\imath}]\|_s \leq_s \varepsilon^3\big(\|\hat{\imath}\|_{s+1} + \|\mathfrak{I}_\delta\|_{s+1}\|\hat{\imath}\|_{s_0+1}\big) \tag{7.38}$$

$$\|p_{\geq 4}\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^4 + \varepsilon^{b+2}\|\mathfrak{I}_\delta\|_{s+2}^{\mathrm{Lip}(\gamma)}, \qquad \|\partial_i p_{\geq 4}[\hat{\imath}]\|_s \leq_s \varepsilon^{b+2}\big(\|\hat{\imath}\|_{s+2} + \|\mathfrak{I}_\delta\|_{s+2}\|\hat{\imath}\|_{s_0+2}\big) \tag{7.39}$$

where $\mathfrak{I}_\delta(\varphi) := (\theta_0(\varphi) - \varphi, y_\delta(\varphi), z_0(\varphi))$ corresponds to T_δ. The remainder $\mathcal{R}_2$ has the form (7.2) with

$$\|g_j\|_s^{\mathrm{Lip}(\gamma)} + \|\chi_j\|_s^{\mathrm{Lip}(\gamma)} \leq_s 1 + \|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \quad \|\partial_i g_j[\hat{\imath}]\|_s + \|\partial_i \chi_j[\hat{\imath}]\|_s \leq_s \|\hat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma} \tag{7.40}$$

and also $\mathcal{R}_*$ has the form (7.2) with

$$\|g_j^*\|_s^{\mathrm{Lip}(\gamma)}\|\chi_j^*\|_{s_0}^{\mathrm{Lip}(\gamma)} + \|g_j^*\|_{s_0}^{\mathrm{Lip}(\gamma)}\|\chi_j^*\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^3 + \varepsilon^{b+1}\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)} \tag{7.41}$$

$$\begin{aligned}&\|\partial_i g_j^*[\hat{\imath}]\|_s\|\chi_j^*\|_{s_0} + \|\partial_i g_j^*[\hat{\imath}]\|_{s_0}\|\chi_j^*\|_s + \|g_j^*\|_{s_0}\|\partial_i \chi_j^*[\hat{\imath}]\|_s + \|g_j^*\|_s\|\partial_i \chi_j^*[\hat{\imath}]\|_{s_0} \\ &\quad \leq_s \varepsilon^{b+1}\|\hat{\imath}\|_{s+\sigma} + \varepsilon^{2b-1}\|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat{\imath}\|_{s_0+\sigma}.\end{aligned} \tag{7.42}$$

The bounds (7.40), (7.41) imply, by Lemma 7.3, estimates for the s-decay norms of $\mathcal{R}_2$ and $\mathcal{R}_*$. The linearized operator $\mathcal{L}_\omega := \mathcal{L}_\omega(\omega, i_\delta(\omega))$ depends on the parameter ω both directly and also through the dependence on the torus $i_\delta(\omega)$. We have estimated also the partial derivative ∂_i with respect to the variables i (see (5.1)) in order to control, along the nonlinear Nash–Moser iteration, the Lipschitz variation of the eigenvalues of $\mathcal{L}_\omega$ with respect to ω and the approximate solution i_δ.

8. Reduction of the linearized operator in the normal directions

The goal of this section is to conjugate the Hamiltonian operator $\mathcal{L}_\omega$ in (7.34) to the diagonal operator $\mathcal{L}_\infty$ defined in (8.121). The proof is obtained applying different kind of symplectic transformations. We shall always assume (7.8).

8.1. Change of the space variable

The first task is to conjugate $\mathcal{L}_\omega$ in (7.34) to $\mathcal{L}_1$ in (8.31), which has the coefficient of ∂_{xxx} independent on the space variable. We look for a φ-dependent family of *symplectic* diffeomorphisms $\Phi(\varphi)$ of $H_S^\perp$ which differ from

$$\mathcal{A}_\perp := \Pi_S^\perp \mathcal{A} \Pi_S^\perp, \quad (\mathcal{A}h)(\varphi, x) := (1 + \beta_x(\varphi, x)) h(\varphi, x + \beta(\varphi, x)), \tag{8.1}$$

up to a small "finite dimensional" remainder, see (8.6). Each $\mathcal{A}(\varphi)$ is a symplectic map of the phase space, see [2]-Remark 3.3. If $\|\beta\|_{W^{1,\infty}} < 1/2$ then $\mathcal{A}$ is invertible, see Lemma 2.4, and its inverse and adjoint maps are

$$(\mathcal{A}^{-1}h)(\varphi, y) := (1+\tilde{\beta}_y(\varphi, y))h(\varphi, y+\tilde{\beta}(\varphi, y)), \quad (\mathcal{A}^T h)(\varphi, y) = h(\varphi, y+\tilde{\beta}(\varphi, y)) \tag{8.2}$$

where $x = y + \tilde{\beta}(\varphi, y)$ is the inverse diffeomorphism (of $\mathbb{T}$) of $y = x + \beta(\varphi, x)$.

The restricted maps $\mathcal{A}_\perp(\varphi) : H_S^\perp \to H_S^\perp$ are not symplectic. In order to find a symplectic diffeomorphism near $\mathcal{A}_\perp(\varphi)$, the first observation is that each $\mathcal{A}(\varphi)$ can be seen as the time 1-flow of a time dependent Hamiltonian PDE. Indeed $\mathcal{A}(\varphi)$ (for simplicity we skip the dependence on φ) is homotopic to the identity via the path of symplectic diffeomorphisms

$$u \mapsto (1+\tau\beta_x)u(x+\tau\beta(x)), \quad \tau \in [0,1],$$

which is the trajectory solution of the time dependent, linear Hamiltonian PDE

$$\partial_\tau u = \partial_x(b(\tau, x)u), \quad b(\tau, x) := \frac{\beta(x)}{1+\tau\beta_x(x)}, \tag{8.3}$$

with value $u(x)$ at $\tau = 0$ and $\mathcal{A}u = (1+\beta_x(x))u(x+\beta(x))$ at $\tau = 1$. The equation (8.3) is a *transport* equation. Its associated characteristic ODE is

$$\frac{d}{d\tau}x = -b(\tau, x). \tag{8.4}$$

We denote its flow by $\gamma^{\tau_0,\tau}$, namely $\gamma^{\tau_0,\tau}(y)$ is the solution of (8.4) with $\gamma^{\tau_0,\tau_0}(y) = y$. Each $\gamma^{\tau_0,\tau}$ is a diffeomorphism of the torus $\mathbb{T}_x$.

Remark 8.1. Let $y \mapsto y + \tilde{\beta}(\tau, y)$ be the inverse diffeomorphism of $x \mapsto x + \tau\beta(x)$. Differentiating the identity $\tilde{\beta}(\tau, y) + \tau\beta(y + \tilde{\beta}(\tau, y)) = 0$ with respect to τ it results that $\gamma^\tau(y) := \gamma^{0,\tau}(y) = y + \tilde{\beta}(\tau, y)$.

Then we define a symplectic map Φ of $H_S^\perp$ as the time-1 flow of the Hamiltonian PDE

$$\partial_\tau u = \Pi_S^\perp \partial_x(b(\tau, x)u) = \partial_x(b(\tau, x)u) - \Pi_S\partial_x(b(\tau, x)u), \quad u \in H_S^\perp. \tag{8.5}$$

Note that $\Pi_S^\perp \partial_x(b(\tau, x)u)$ is the Hamiltonian vector field generated by $\frac{1}{2}\int_{\mathbb{T}} b(\tau, x)u^2 dx$ restricted to $H_S^\perp$. We denote by $\Phi^{\tau_0,\tau}$ the flow of (8.5), namely $\Phi^{\tau_0,\tau}(u_0)$ is the solution of (8.5) with initial condition $\Phi^{\tau_0,\tau_0}(u_0) = u_0$. The flow is well defined in Sobolev spaces $H^s_{S^\perp}(\mathbb{T}_x)$ for $b(\tau, x)$ smooth enough (standard theory of linear hyperbolic PDEs, see e.g. Section 0.8 in [31]). It is natural to expect that the difference between the flow map $\Phi := \Phi^{0,1}$ and $\mathcal{A}_\perp$ is a "finite-dimensional" remainder of the size of β.

Lemma 8.2. *For $\|\beta\|_{W^{s_0+1,\infty}}$ small, there exists an invertible symplectic transformation $\Phi = \mathcal{A}_\perp + \mathcal{R}_\Phi$ of $H^s_{S^\perp}$, where $\mathcal{A}_\perp$ is defined in (8.1) and $\mathcal{R}_\Phi$ is a "finite-dimensional" remainder*

$$\mathcal{R}_\Phi h = \sum_{j\in S}\int_0^1 (h, g_j(\tau))_{L^2(\mathbb{T})}\chi_j(\tau)d\tau + \sum_{j\in S}\big(h, \psi_j\big)_{L^2(\mathbb{T})}e^{\mathrm{i}jx} \tag{8.6}$$

for some functions $\chi_j(\tau), g_j(\tau), \psi_j \in H^s$ satisfying

$$\|\psi_j\|_s, \|g_j(\tau)\|_s \leq_s \|\beta\|_{W^{s+2,\infty}}, \quad \|\chi_j(\tau)\|_s \leq_s 1 + \|\beta\|_{W^{s+1,\infty}}, \quad \forall \tau \in [0,1]. \tag{8.7}$$

Furthermore, the following tame estimates holds

$$\|\Phi^{\pm 1}h\|_s \leq_s \|h\|_s + \|\beta\|_{W^{s+2,\infty}}\|h\|_{s_0}, \quad \forall h \in H^s_{S^\perp}. \tag{8.8}$$

Proof. Let $w(\tau, x) := (\Phi^\tau u_0)(x)$ denote the solution of (8.5) with initial condition $\Phi^0(w) = u_0 \in H_S^\perp$. The difference

$$(\mathcal{A}_\perp - \Phi)u_0 = \Pi_S^\perp \mathcal{A}u_0 - w(1,\cdot) = \mathcal{A}u_0 - w(1,\cdot) - \Pi_S\mathcal{A}u_0, \quad \forall u_0 \in H_S^\perp, \tag{8.9}$$

and

$$\Pi_S \mathcal{A} u_0 = \Pi_S(\mathcal{A} - I)\Pi_S^\perp u_0 = \sum_{j\in S} \big(u_0\,,\, \psi_j\big)_{L^2(\mathbb{T})} e^{\mathrm{i}jx}\,, \quad \psi_j := (\mathcal{A}^T - I)e^{\mathrm{i}jx}\,. \tag{8.10}$$

We claim that the difference

$$\mathcal{A}u_0 - w(1,x) = (1+\beta_x(x)) \int_0^1 (1+\tau\beta_x(x))^{-1}\big[\Pi_S \partial_x(b(\tau)w(\tau))\big](\gamma^\tau(x+\beta(x)))\,d\tau \tag{8.11}$$

where $\gamma^\tau(y) := \gamma^{0,\tau}(y)$ is the flow of (8.4). Indeed the solution $w(\tau,x)$ of (8.5) satisfies

$$\partial_\tau\{w(\tau,\gamma^\tau(y))\} = b_x(\tau,\gamma^\tau(y))w(\tau,\gamma^\tau(y)) - \big[\Pi_S\partial_x(b(\tau)w(\tau))\big](\gamma^\tau(y))\,.$$

Then, by the variation of constant formula, we find

$$w(\tau,\gamma^\tau(y)) = e^{\int_0^\tau b_x(s,\gamma^s(y))\,ds}\Big(u_0(y) - \int_0^\tau e^{-\int_0^s b_x(\zeta,\gamma^\zeta(y))\,d\zeta}\big[\Pi_S\partial_x(b(s)w(s))\big](\gamma^s(y))\,ds\Big)\,.$$

Since $\partial_y\gamma^\tau(y)$ solves the variational equation $\partial_\tau(\partial_y\gamma^\tau(y)) = -b_x(\tau,\gamma^\tau(y))(\partial_y\gamma^\tau(y))$ with $\partial_y\gamma^0(y) = 1$ we have that

$$e^{\int_0^\tau b_x(s,\gamma^s(y))ds} = \big(\partial_y\gamma^\tau(y)\big)^{-1} = 1+\tau\beta_x(x) \tag{8.12}$$

by Remark 8.1, and so we derive the expression

$$w(\tau,x) = (1+\tau\beta_x(x))\Big\{u_0(x+\tau\beta(x)) - \int_0^\tau (1+s\beta_x(x))^{-1}\big[\Pi_S\partial_x(b(s)w(s))\big](\gamma^s(x+\tau\beta(x)))\,ds\Big\}\,.$$

Evaluating at $\tau = 1$, formula (8.11) follows. Next, we develop (recall $w(\tau) = \Phi^\tau(u_0)$)

$$[\Pi_S\partial_x(b(\tau)w(\tau))](x) = \sum_{j\in S}\big(u_0, g_j(\tau)\big)_{L^2(\mathbb{T})} e^{\mathrm{i}jx}\,, \quad g_j(\tau) := -(\Phi^\tau)^T[b(\tau)\partial_x e^{\mathrm{i}jx}]\,, \tag{8.13}$$

and (8.11) becomes

$$\mathcal{A}u_0 - w(1,\cdot) = -\int_0^1 \sum_{j\in S}\big(u_0\,,\, g_j(\tau)\big)_{L^2(\mathbb{T})}\chi_j(\tau,\cdot)\,d\tau\,, \tag{8.14}$$

where

$$\chi_j(\tau,x) := -(1+\beta_x(x))(1+\tau\beta_x(x))^{-1}e^{\mathrm{i}j\gamma^\tau(x+\beta(x))}\,. \tag{8.15}$$

By (8.9), (8.10), (8.11), (8.14) we deduce that $\Phi = \mathcal{A}_\perp + \mathcal{R}_\Phi$ as in (8.6).

We now prove the estimates (8.7). Each function ψ_j in (8.10) satisfies $\|\psi_j\|_s \le_s \|\beta\|_{W^{s,\infty}}$, see (8.2). The bound $\|\chi_j(\tau)\|_s \le_s 1 + \|\beta\|_{W^{s+1,\infty}}$ follows by (8.15). The tame estimates for $g_j(\tau)$ defined in (8.13) are more difficult because require tame estimates for the adjoint $(\Phi^\tau)^T$, $\forall\tau\in[0,1]$. The adjoint of the flow map can be represented as the flow map of the "adjoint" PDE

$$\partial_\tau z = \Pi_S^\perp\{b(\tau,x)\partial_x\Pi_S^\perp z\} = b(\tau,x)\partial_x z - \Pi_S(b(\tau,x)\partial_x z)\,, \quad z\in H_S^\perp\,, \tag{8.16}$$

where $-\Pi_S^\perp b(\tau,x)\partial_x$ is the L^2-adjoint of the Hamiltonian vector field in (8.5). We denote by $\Psi^{\tau_0,\tau}$ the flow of (8.16), namely $\Psi^{\tau_0,\tau}(v)$ is the solution of (8.16) with $\Psi^{\tau_0,\tau_0}(v) = v$. Since the derivative $\partial_\tau(\Phi^\tau(u_0), \Psi^{\tau_0,\tau}(v))_{L^2(\mathbb{T})} = 0$, $\forall\tau$, we deduce that $(\Phi^{\tau_0}(u_0), \Psi^{\tau_0,\tau_0}(v))_{L^2(\mathbb{T})} = (\Phi^0(u_0), \Psi^{\tau_0,0}(v))_{L^2(\mathbb{T})}$, namely (recall that $\Psi^{\tau_0,\tau_0}(v) = v$) the adjoint

$$(\Phi^{\tau_0})^T = \Psi^{\tau_0,0}\,, \quad \forall\tau_0\in[0,1]\,. \tag{8.17}$$

Thus it is sufficient to prove tame estimates for the flow $\Psi^{\tau_0,\tau}$. We first provide a useful expression for the solution $z(\tau,x) := \Psi^{\tau_0,\tau}(v)$ of (8.16), obtained by the methods of characteristics. Let $\gamma^{\tau_0,\tau}(y)$ be the flow of (8.4). Since $\partial_\tau z(\tau,\gamma^{\tau_0,\tau}(y)) = -[\Pi_S(b(\tau)\partial_x z(\tau)](\gamma^{\tau_0,\tau}(y))$ we get

$$z(\tau,\gamma^{\tau_0,\tau}(y)) = v(y) + \int_\tau^{\tau_0} [\Pi_S(b(s)\partial_x z(s)](\gamma^{\tau_0,s}(y))\,ds\,, \quad \forall \tau \in [0,1]\,.$$

Denoting by $y = x + \sigma(\tau,x)$ the inverse diffeomorphism of $x = \gamma^{\tau_0,\tau}(y) = y + \tilde\sigma(\tau,y)$, we get

$$\begin{aligned}\Psi^{\tau_0,\tau}(v) = z(\tau,x) &= v(x+\sigma(\tau,x)) + \int_\tau^{\tau_0} [\Pi_S(b(s)\partial_x z(s)](\gamma^{\tau_0,s}(x+\sigma(\tau,x)))\,ds \\ &= v(x+\sigma(\tau,x)) + \int_\tau^{\tau_0} \sum_{j\in S}(z(s),p_j(s))\kappa_j(s,x)\,ds = v(x+\sigma(\tau,x)) + \mathcal{R}_\tau v\,,\end{aligned} \tag{8.18}$$

where $p_j(s) := -\partial_x(b(s)e^{\mathrm{i}jx})$, $\kappa_j(s,x) := e^{\mathrm{i}j\gamma^{\tau_0,s}(x+\sigma(\tau,x))}$ and

$$(\mathcal{R}_\tau v)(x) := \int_\tau^{\tau_0} \sum_{j\in S}(\Psi^{\tau_0,s}(v),p_j(s))_{L^2(\mathbb{T})}\kappa_j(s,x)\,ds\,.$$

Since $\|\sigma(\tau,\cdot)\|_{W^{s,\infty}}, \|\tilde\sigma(\tau,\cdot)\|_{W^{s,\infty}} \le_s \|\beta\|_{W^{s+1,\infty}}$ (recall also (8.3)), we derive $\|p_j\|_s \le_s \|\beta\|_{W^{s+2,\infty}}$, $\|\kappa_j\|_s \le_s 1 + \|\beta\|_{W^{s+1,\infty}}$ and $\|v(x+\sigma(\tau,x))\|_s \le_s \|v\|_s + \|\beta\|_{W^{s+1,\infty}}\|v\|_{s_0}$, $\forall\tau\in[0,1]$. Moreover

$$\|\mathcal{R}_\tau v\|_s \le_s \sup_{\tau\in[0,1]}\|\Psi^{\tau_0,\tau}(v)\|_s\|\beta\|_{W^{s_0+2,\infty}} + \sup_{\tau\in[0,1]}\|\Psi^{\tau_0,\tau}(v)\|_{s_0}\|\beta\|_{W^{s+2,\infty}}\,.$$

Therefore, for all $\tau\in[0,1]$,

$$\|\Psi^{\tau_0,\tau}v\|_s \le_s \|v\|_s + \|\beta\|_{W^{s+1,\infty}}\|v\|_{s_0} + \sup_{\tau\in[0,1]}\big\{\|\Psi^{\tau_0,\tau}v\|_s\|\beta\|_{W^{s_0+2,\infty}} + \|\Psi^{\tau_0,\tau}v\|_{s_0}\|\beta\|_{W^{s+2,\infty}}\big\}\,. \tag{8.19}$$

For $s = s_0$ it implies

$$\sup_{\tau\in[0,1]}\|\Psi^{\tau_0,\tau}(v)\|_{s_0} \le_{s_0} \|v\|_{s_0}(1+\|\beta\|_{W^{s_0+1,\infty}}) + \sup_{\tau\in[0,1]}\|\Psi^{\tau_0,\tau}(v)\|_{s_0}\|\beta\|_{W^{s_0+2,\infty}}$$

and so, for $\|\beta\|_{W^{s_0+2,\infty}} \le c(s_0)$ small enough,

$$\sup_{\tau\in[0,1]}\|\Psi^{\tau_0,\tau}(v)\|_{s_0} \le_{s_0} \|v\|_{s_0}\,. \tag{8.20}$$

Finally (8.19), (8.20) imply the tame estimate

$$\sup_{\tau\in[0,1]}\|\Psi^{\tau_0,\tau}(v)\|_s \le_s \|v\|_s + \|\beta\|_{W^{s+2,\infty}}\|v\|_{s_0}\,. \tag{8.21}$$

By (8.17) and (8.21) we deduce the bound (8.7) for g_j defined in (8.13). The tame estimate (8.8) for Φ follows by that of $\mathcal{A}$ and (8.7) (use Lemma 2.4). The estimate for Φ^{-1} follows in the same way because $\Phi^{-1} = \Phi^{1,0}$ is the backward flow. □

We conjugate $\mathcal{L}_\omega$ in (7.34) via the symplectic map $\Phi = \mathcal{A}_\perp + \mathcal{R}_\Phi$ of Lemma 8.2. We compute (split $\Pi_S^\perp = I - \Pi_S$)

$$\mathcal{L}_\omega\Phi = \Phi\mathcal{D}_\omega + \Pi_S^\perp\mathcal{A}\big(b_3\partial_{yyy} + b_2\partial_{yy} + b_1\partial_y + b_0\big)\Pi_S^\perp + \mathcal{R}_I\,, \tag{8.22}$$

where the coefficients are

$$b_3(\varphi,y) := \mathcal{A}^T[a_1(1+\beta_x)^3] \qquad b_2(\varphi,y) := \mathcal{A}^T\big[2(a_1)_x(1+\beta_x)^2 + 6a_1\beta_{xx}(1+\beta_x)\big] \tag{8.23}$$

$$b_1(\varphi,y) := \mathcal{A}^T\Big[(\mathcal{D}_\omega\beta) + 3a_1\frac{\beta_{xx}^2}{1+\beta_x} + 4a_1\beta_{xxx} + 6(a_1)_x\beta_{xx} + (a_1)_{xx}(1+\beta_x) + a_0(1+\beta_x)\Big] \tag{8.24}$$

$$b_0(\varphi,y) := \mathcal{A}^T\Big[\frac{(\mathcal{D}_\omega\beta_x)}{1+\beta_x} + a_1\frac{\beta_{xxxx}}{1+\beta_x} + 2(a_1)_x\frac{\beta_{xxx}}{1+\beta_x} + (a_1)_{xx}\frac{\beta_{xx}}{1+\beta_x} + a_0\frac{\beta_{xx}}{1+\beta_x} + (a_0)_x\Big] \tag{8.25}$$

and the remainder

$$\mathcal{R}_I := -\Pi_S^\perp \partial_x(\varepsilon^2\mathcal{R}_2 + \mathcal{R}_*)\mathcal{A}_\perp - \Pi_S^\perp\big(a_1\partial_{xxx} + 2(a_1)_x\partial_{xx} + ((a_1)_{xx} + a_0)\partial_x + (a_0)_x\big)\Pi_S\mathcal{A}\Pi_S^\perp + [\mathcal{D}_\omega, \mathcal{R}_\Phi] + (\mathcal{L}_\omega - \mathcal{D}_\omega)\mathcal{R}_\Phi\,. \tag{8.26}$$

The commutator $[\mathcal{D}_\omega, \mathcal{R}_\Phi]$ has the form (8.6) with $\mathcal{D}_\omega g_j$ or $\mathcal{D}_\omega\chi_j$, $\mathcal{D}_\omega\psi_j$ instead of χ_j, g_j, ψ_j respectively. Also the last term $(\mathcal{L}_\omega - \mathcal{D}_\omega)\mathcal{R}_\Phi$ in (8.26) has the form (8.6) (note that $\mathcal{L}_\omega - \mathcal{D}_\omega$ does not contain derivatives with respect to φ). By (8.22), and decomposing $I = \Pi_S + \Pi_S^\perp$, we get

$$\mathcal{L}_\omega\Phi = \Phi(\mathcal{D}_\omega + b_3\partial_{yyy} + b_2\partial_{yy} + b_1\partial_y + b_0)\Pi_S^\perp + \mathcal{R}_{II}\,, \tag{8.27}$$

$$\mathcal{R}_{II} := \big\{\Pi_S^\perp(\mathcal{A} - I)\Pi_S - \mathcal{R}_\Phi\big\}(b_3\partial_{yyy} + b_2\partial_{yy} + b_1\partial_y + b_0)\Pi_S^\perp + \mathcal{R}_I\,. \tag{8.28}$$

Now we choose the function $\beta = \beta(\varphi, x)$ such that

$$a_1(\varphi, x)(1 + \beta_x(\varphi, x))^3 = b_3(\varphi) \tag{8.29}$$

so that the coefficient b_3 in (8.23) depends only on φ (note that $\mathcal{A}^T[b_3(\varphi)] = b_3(\varphi)$). The only solution of (8.29) with zero space average is (see e.g. [2]-Section 3.1)

$$\beta := \partial_x^{-1}\rho_0, \quad \rho_0 := b_3(\varphi)^{1/3}(a_1(\varphi, x))^{-1/3} - 1, \quad b_3(\varphi) := \Big(\frac{1}{2\pi}\int_{\mathbb{T}}(a_1(\varphi, x))^{-1/3}dx\Big)^{-3}. \tag{8.30}$$

Applying the symplectic map Φ^{-1} in (8.27) we obtain the Hamiltonian operator (see Definition 2.2)

$$\mathcal{L}_1 := \Phi^{-1}\mathcal{L}_\omega\Phi = \Pi_S^\perp\big(\omega\cdot\partial_\varphi + b_3(\varphi)\partial_{yyy} + b_1\partial_y + b_0\big)\Pi_S^\perp + \mathfrak{R}_1 \tag{8.31}$$

where $\mathfrak{R}_1 := \Phi^{-1}\mathcal{R}_{II}$. We used that, by the Hamiltonian nature of $\mathcal{L}_1$, the coefficient $b_2 = 2(b_3)_y$ (see [2]-Remark 3.5) and so, by the choice (8.30), we have $b_2 = 2(b_3)_y = 0$. In the next lemma we analyze the structure of the remainder $\mathfrak{R}_1$.

Lemma 8.3. *The operator $\mathfrak{R}_1$ has the form (7.7).*

Proof. The remainders $\mathcal{R}_I$ and $\mathcal{R}_{II}$ have the form (7.7). Indeed $\mathcal{R}_2$, $\mathcal{R}_*$ in (8.26) have the form (7.2) (see Proposition 7.6) and the term $\Pi_S\mathcal{A}w = \sum_{j\in S}(\mathcal{A}^T e^{\mathrm{i}jx}, w)_{L^2(\mathbb{T})}e^{\mathrm{i}jx}$ has the same form. By (8.6), the terms of $\mathcal{R}_I$, $\mathcal{R}_{II}$ which involves the operator $\mathcal{R}_\Phi$ have the form (7.7). All the operations involved preserve this structure: if $R_\tau w = \chi(\tau)(w, g(\tau))_{L^2(\mathbb{T})}$, $\tau\in[0,1]$, then

$$R_\tau\Pi_S^\perp w = \chi(\tau)(\Pi_S^\perp g(\tau), w)_{L^2(\mathbb{T})}\,, \quad R_\tau\mathcal{A}w = \chi(\tau)(\mathcal{A}^T g(\tau), w)_{L^2(\mathbb{T})}\,, \quad \partial_x R_\tau w = \chi_x(\tau)(g(\tau), w)_{L^2(\mathbb{T})}\,,$$
$$\Pi_S^\perp R_\tau w = (\Pi_S^\perp\chi(\tau))(g(\tau), w)_{L^2(\mathbb{T})}\,, \quad \mathcal{A}R_\tau w = (\mathcal{A}\chi(\tau))(g(\tau), w)_{L^2(\mathbb{T})}\,,$$
$$\Phi^{-1}R_\tau w = (\Phi^{-1}\chi(\tau))(g(\tau), w)_{L^2(\mathbb{T})}$$

(the last equality holds because $\Phi^{-1}(f(\varphi)w) = f(\varphi)\Phi^{-1}(w)$ for all function $f(\varphi)$). Hence $\mathfrak{R}_1$ has the form (7.7) where $\chi_j(\tau)\in H_S^\perp$ for all $\tau\in[0,1]$. □

We now put in evidence the terms of order $\varepsilon, \varepsilon^2, \ldots,$ in b_1, b_0, $\mathfrak{R}_1$, recalling that $a_1 - 1 = O(\varepsilon^3)$ (see (7.38)), $a_0 = O(\varepsilon)$ (see (7.35)–(7.39)), and $\beta = O(\varepsilon^3)$ (proved below in (8.35)). We expand b_1 in (8.24) as

$$b_1 = -\varepsilon p_1 - \varepsilon^2 p_2 - q_{>2} + \mathcal{D}_\omega\beta + 4\beta_{xxx} + (a_1)_{xx} + b_{1,\geq 4} \tag{8.32}$$

where $b_{1,\geq 4} = O(\varepsilon^4)$ is defined by difference (the precise estimate is in Lemma 8.5).

Remark 8.4. The function $\mathcal{D}_\omega \beta$ has zero average in x by (8.30) as well as $(a_1)_{xx}$, β_{xxx}.

Similarly, we expand b_0 in (8.25) as

$$b_0 = -\varepsilon(p_1)_x - \varepsilon^2(p_2)_x - (q_{>2})_x + \mathcal{D}_\omega \beta_x + \beta_{xxxx} + b_{0,\geq 4} \tag{8.33}$$

where $b_{0,\geq 4} = O(\varepsilon^4)$ is defined by difference.

Using the equalities (8.28), (8.26) and $\Pi_S \mathcal{A} \Pi_S^\perp = \Pi_S(\mathcal{A} - I)\Pi_S^\perp$ we get

$$\mathfrak{R}_1 := \Phi^{-1}\mathcal{R}_{II} = -\varepsilon^2 \Pi_S^\perp \partial_x \mathcal{R}_2 + \mathcal{R}_* \tag{8.34}$$

where $\mathcal{R}_2$ is defined in (7.29) and we have renamed $\mathcal{R}_*$ the term of order $o(\varepsilon^2)$ in $\mathfrak{R}_1$. The remainder $\mathcal{R}_*$ in (8.34) has the form (7.7).

Lemma 8.5. *There is $\sigma = \sigma(\tau, \nu) > 0$ such that*

$$\|\beta\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^3(1 + \|\mathfrak{I}_\delta\|_{s+1}^{\mathrm{Lip}(\gamma)}), \qquad \|\partial_i \beta[\widehat{\imath}]\|_s \leq_s \varepsilon^3\big(\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\big), \tag{8.35}$$

$$\|b_3 - 1\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^4 + \varepsilon^{b+2}\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \qquad \|\partial_i b_3[\widehat{\imath}]\|_s \leq_s \varepsilon^{b+2}\big(\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\big) \tag{8.36}$$

$$\|b_{1,\geq 4}\|_s^{\mathrm{Lip}(\gamma)} + \|b_{0,\geq 4}\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^4 + \varepsilon^{b+2}\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)} \tag{8.37}$$

$$\|\partial_i b_{1,\geq 4}[\widehat{\imath}]\|_s + \|\partial_i b_{0,\geq 4}[\widehat{\imath}]\|_s \leq_s \varepsilon^{b+2}\big(\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\big). \tag{8.38}$$

The transformations Φ, Φ^{-1} satisfy

$$\|\Phi^{\pm 1} h\|_s^{\mathrm{Lip}(\gamma)} \leq_s \|h\|_{s+1}^{\mathrm{Lip}(\gamma)} + \|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}\|h\|_{s_0+1}^{\mathrm{Lip}(\gamma)} \tag{8.39}$$

$$\|\partial_i(\Phi^{\pm 1} h)[\widehat{\imath}]\|_s \leq_s \|h\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma} + \|h\|_{s_0+\sigma}\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|h\|_{s_0+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\,. \tag{8.40}$$

Moreover the remainder $\mathcal{R}_*$ has the form (7.7), where the functions $\chi_j(\tau)$, $g_j(\tau)$ satisfy the estimates (7.41)–(7.42) uniformly in $\tau \in [0, 1]$.

Proof. The estimates (8.35) follow by (8.30), (7.38), and the usual interpolation and tame estimates in Lemmata 2.2–2.4 (and Lemma 5.13) and (7.8). For the estimates of b_3, by (8.30) and (7.35) we consider the function r_1 defined in (7.25). Recalling also (3.4) and (7.11), the function

$$r_1(T_\delta) = \varepsilon^3(\partial_{u_x u_x} f_5)(v_\delta, (v_\delta)_x) + r_{1,\geq 4}\,, \quad r_{1,\geq 4} := r_1(T_\delta) - \varepsilon^3(\partial_{u_x u_x} f_5)(v_\delta, (v_\delta)_x)\,.$$

Hypothesis (S1) implies, as in the proof of Lemma 7.5, that the space average $\int_{\mathbb{T}}(\partial_{u_x u_x} f_5)(v_\delta, (v_\delta)_x)dx = 0$. Hence the bound (8.36) for $b_3 - 1$ follows. For the estimates on Φ, Φ^{-1} we apply Lemma 8.2 and the estimate (8.35) for β. We estimate the remainder $\mathcal{R}_*$ in (8.34), using (8.26), (8.28) and (7.41)–(7.42). $\square$

8.2. Reparametrization of time

The goal of this section is to make constant the coefficient of the highest order spatial derivative operator ∂_{yyy}, by a quasi-periodic reparametrization of time. We consider the change of variable

$$(Bw)(\varphi, y) := w(\varphi + \omega\alpha(\varphi), y), \qquad (B^{-1}h)(\vartheta, y) := h(\vartheta + \omega\tilde{\alpha}(\vartheta), y)\,,$$

where $\varphi = \vartheta + \omega\tilde{\alpha}(\vartheta)$ is the inverse diffeomorphism of $\vartheta = \varphi + \omega\alpha(\varphi)$ in $\mathbb{T}^\nu$. By conjugation, the differential operators become

$$B^{-1}\omega \cdot \partial_\varphi B = \rho(\vartheta)\,\omega \cdot \partial_\vartheta\,, \quad B^{-1}\partial_y B = \partial_y\,, \quad \rho := B^{-1}(1 + \omega \cdot \partial_\varphi \alpha). \tag{8.41}$$

By (8.31), using also that B and B^{-1} commute with $\Pi_S^\perp$, we get

$$B^{-1}\mathcal{L}_1 B = \Pi_S^\perp[\rho\omega \cdot \partial_\vartheta + (B^{-1}b_3)\partial_{yyy} + (B^{-1}b_1)\partial_y + (B^{-1}b_0)]\Pi_S^\perp + B^{-1}\mathfrak{R}_1 B. \tag{8.42}$$

We choose α such that

$$(B^{-1}b_3)(\vartheta) = m_3\rho(\vartheta), \quad m_3 \in \mathbb{R}, \quad \text{namely} \quad b_3(\varphi) = m_3(1 + \omega \cdot \partial_\varphi \alpha(\varphi)) \tag{8.43}$$

(recall (8.41)). The unique solution with zero average of (8.43) is

$$\alpha(\varphi) := \frac{1}{m_3}(\omega \cdot \partial_\varphi)^{-1}(b_3 - m_3)(\varphi), \qquad m_3 := \frac{1}{(2\pi)^\nu}\int_{\mathbb{T}^\nu} b_3(\varphi)d\varphi\,. \tag{8.44}$$

Hence, by (8.42),

$$B^{-1}\mathcal{L}_1 B = \rho\mathcal{L}_2\,, \qquad \mathcal{L}_2 := \Pi_S^\perp(\omega \cdot \partial_\vartheta + m_3\partial_{yyy} + c_1\partial_y + c_0)\Pi_S^\perp + \mathfrak{R}_2 \tag{8.45}$$

$$c_1 := \rho^{-1}(B^{-1}b_1)\,, \qquad c_0 := \rho^{-1}(B^{-1}b_0)\,, \qquad \mathfrak{R}_2 := \rho^{-1}B^{-1}\mathfrak{R}_1 B\,. \tag{8.46}$$

The transformed operator $\mathcal{L}_2$ in (8.45) is still Hamiltonian, since the reparametrization of time preserves the Hamiltonian structure (see Section 2.2 and Remark 3.7 in [2]).

We now put in evidence the terms of order $\varepsilon, \varepsilon^2, \ldots$ in c_1, c_0. To this aim, we anticipate the following estimates: $\rho(\vartheta) = 1 + O(\varepsilon^4)$, $\alpha = O(\varepsilon^4\gamma^{-1})$, $m_3 = 1 + O(\varepsilon^4)$, $B^{-1} - I = O(\alpha)$ (in low norm), which are proved in Lemma 8.7 below. Then, by (8.32)–(8.33), we expand the functions c_1, c_0 in (8.46) as

$$c_1 = -\varepsilon p_1 - \varepsilon^2 p_2 - B^{-1}q_{>2} + \varepsilon(p_1 - B^{-1}p_1) + \varepsilon^2(p_2 - B^{-1}p_2) + \mathcal{D}_\omega\beta + 4\beta_{xxx} + (a_1)_{xx} + c_{1,\geq 4}\,, \tag{8.47}$$

$$c_0 = -\varepsilon(p_1)_x - \varepsilon^2(p_2)_x - (B^{-1}q_{>2})_x + \varepsilon(p_1 - B^{-1}p_1)_x + \varepsilon^2(p_2 - B^{-1}p_2)_x + (\mathcal{D}_\omega\beta)_x + \beta_{xxxx} + c_{0,\geq 4}\,, \tag{8.48}$$

where $c_{1,\geq 4}, c_{0,\geq 4} = O(\varepsilon^4)$ are defined by difference.

Remark 8.6. The functions $\varepsilon(p_1 - B^{-1}p_1) = O(\varepsilon^5\gamma^{-1})$ and $\varepsilon^2(p_2 - B^{-1}p_2) = O(\varepsilon^6\gamma^{-1})$, see (8.53). For the reducibility scheme, the terms of order ∂_x^0 with size $O(\varepsilon^5\gamma^{-1})$ are perturbative, since $\varepsilon^5\gamma^{-2} \ll 1$.

The remainder $\mathfrak{R}_2$ in (8.46) has still the form (7.7) and, by (8.34),

$$\mathfrak{R}_2 := -\rho^{-1}B^{-1}\mathfrak{R}_1 B = -\varepsilon^2\Pi_S^\perp\partial_x\mathcal{R}_2 + \mathcal{R}_* \tag{8.49}$$

where $\mathcal{R}_2$ is defined in (7.29) and we have renamed $\mathcal{R}_*$ the term of order $o(\varepsilon^2)$ in $\mathfrak{R}_2$.

Lemma 8.7. *There is $\sigma = \sigma(\nu, \tau) > 0$ (possibly larger than σ in Lemma 8.5) such that*

$$|m_3 - 1|^{\mathrm{Lip}(\gamma)} \leq C\varepsilon^4, \qquad |\partial_i m_3[\widehat{\imath}]| \leq C\varepsilon^{b+2}\|\widehat{\imath}\|_{s_0+\sigma} \tag{8.50}$$

$$\|\alpha\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^4\gamma^{-1} + \varepsilon^{b+2}\gamma^{-1}\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \qquad \|\partial_i\alpha[\widehat{\imath}]\|_s \leq_s \varepsilon^{b+2}\gamma^{-1}\big(\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\big)\,, \tag{8.51}$$

$$\|\rho - 1\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^4 + \varepsilon^{b+2}\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \qquad \|\partial_i\rho[\widehat{\imath}]\|_s \leq_s \varepsilon^{b+2}\big(\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\big) \tag{8.52}$$

$$\|p_k - B^{-1}p_k\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^4\gamma^{-1} + \varepsilon^{b+2}\gamma^{-1}\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \quad k = 1, 2 \tag{8.53}$$

$$\|\partial_i(p_k - B^{-1}p_k)[\widehat{\imath}]\|_s \leq_s \varepsilon^{b+2}\gamma^{-1}\big(\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\big) \tag{8.54}$$

$$\|B^{-1}q_{>2}\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^3 + \varepsilon^b\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \tag{8.55}$$

$$\|\partial_i(B^{-1}q_{>2})[\widehat{\imath}]\|_s \leq_s \varepsilon^b\big(\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\big)\,. \tag{8.56}$$

The terms $c_{1,\geq 4}$, $c_{0,\geq 4}$ satisfy the bounds (8.37)–(8.38). The transformations B, B^{-1} satisfy the estimates (8.39), (8.40). The remainder $\mathcal{R}_*$ has the form (7.7), and the functions $g_j(\tau)$, $\chi_j(\tau)$ satisfy the estimates (7.41)–(7.42) for all $\tau \in [0, 1]$.

Proof. (8.50) follows from (8.44), (8.36). The estimate $\|\alpha\|_s \leq_s \varepsilon^4\gamma^{-1} + \varepsilon^{b+2}\gamma^{-1}\|\mathfrak{I}_\delta\|_{s+\sigma}$ and the inequality for $\partial_i\alpha$ in (8.51) follow by (8.44), (8.36), (8.50). For the first bound in (8.51) we also differentiate (8.44) with respect to the parameter ω. The estimates for ρ follow from $\rho - 1 = B^{-1}(b_3 - m_3)/m_3$. $\square$

8.3. Translation of the space variable

In view of the next linear Birkhoff normal form steps (whose goal is to eliminate the terms of size ε and ε^2), in the expressions (8.47), (8.48) we split $p_1 = \bar p_1 + (p_1 - \bar p_1)$, $p_2 = \bar p_2 + (p_2 - \bar p_2)$ (see (7.35)), where

$$\bar p_1 := 6\bar v, \qquad \bar p_2 := 6\pi_0[(\partial_x^{-1}\bar v)^2], \qquad \bar v(\varphi, x) := \sum_{j\in S} \sqrt{\xi_j} e^{\mathrm{i}\ell(j)\cdot\varphi} e^{\mathrm{i}jx}, \tag{8.57}$$

and $\ell : S \to \mathbb{Z}^\nu$ is the odd injective map (see (1.8))

$$\ell : S \to \mathbb{Z}^\nu, \quad \ell(\bar\jmath_i) := e_i, \quad \ell(-\bar\jmath_i) := -\ell(\bar\jmath_i) = -e_i, \quad i = 1, \dots, \nu, \tag{8.58}$$

denoting by $e_i = (0, \dots, 1, \dots, 0)$ the i-th vector of the canonical basis of $\mathbb{R}^\nu$.

Remark 8.8. All the functions $\bar p_1$, $\bar p_2$, $p_1 - \bar p_1$, $p_2 - \bar p_2$ have zero average in x.

We write the variable coefficients c_1, c_0 of the operator $\mathcal{L}_2$ in (8.45) (see (8.47), (8.48)) as

$$c_1 = -\varepsilon\bar p_1 - \varepsilon^2 \bar p_2 + q_{c_1} + c_{1,\geq 4}\,, \qquad c_0 = -\varepsilon(\bar p_1)_x - \varepsilon^2(\bar p_2)_x + q_{c_0} + c_{0,\geq 4}\,, \tag{8.59}$$

where we define

$$q_{c_1} := q + 4\beta_{xxx} + (a_1)_{xx}\,, \quad q_{c_0} := q_x + \beta_{xxxx}\,, \tag{8.60}$$

$$q := \varepsilon(p_1 - B^{-1}p_1) + \varepsilon(\bar p_1 - p_1) + \varepsilon^2(p_2 - B^{-1}p_2) + \varepsilon^2(\bar p_2 - p_2) - B^{-1}q_{>2} + \mathcal{D}_\omega\beta\,. \tag{8.61}$$

Remark 8.9. The functions q_{c_1}, q_{c_0} have zero average in x (see Remarks 8.8, 8.4 and Lemma 7.5).

Lemma 8.10. *The functions $\bar p_k - p_k$, $k = 1, 2$ and q_{c_m}, $m = 0, 1$, satisfy*

$$\|\bar p_k - p_k\|_s^{\mathrm{Lip}(\gamma)} \leq_s \|\mathfrak{I}_\delta\|_s^{\mathrm{Lip}(\gamma)}, \quad \|\partial_i(\bar p_k - p_k)[\hat\imath]\|_s \leq_s \|\hat\imath\|_s + \|\mathfrak{I}_\delta\|_s\|\hat\imath\|_{s_0}\,, \tag{8.62}$$

$$\|q_{c_m}\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^5\gamma^{-1} + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \quad \|\partial_i q_{c_m}[\hat\imath]\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon\big(\|\hat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\hat\imath\|_{s_0+\sigma}\big)\,. \tag{8.63}$$

Proof. The bound (8.62) follows from (8.57), (7.35), (7.11), (7.8). Then use (8.62), (8.53)–(8.56), (8.35), (7.38) to prove (8.63). The biggest term comes from $\varepsilon(\bar p_1 - p_1)$. □

We now apply the transformation $\mathcal{T}$ defined in (8.64) whose goal is to remove the space average from the coefficient in front of ∂_y.

Consider the change of the space variable $z = y + p(\vartheta)$ which induces on $H^s_{S^\perp}(\mathbb{T}^{\nu+1})$ the operators

$$(\mathcal{T}w)(\vartheta, y) := w(\vartheta, y + p(\vartheta))\,, \quad (\mathcal{T}^{-1}h)(\vartheta, z) = h(\vartheta, z - p(\vartheta)) \tag{8.64}$$

(which are a particular case of those used in Section 8.1). The differential operator becomes $\mathcal{T}^{-1}\omega\cdot\partial_\vartheta\mathcal{T} = \omega\cdot\partial_\vartheta + \{\omega\cdot\partial_\vartheta p(\vartheta)\}\partial_z$, $\mathcal{T}^{-1}\partial_y\mathcal{T} = \partial_z$. Since $\mathcal{T}$, $\mathcal{T}^{-1}$ commute with $\Pi_S^\perp$, we get

$$\mathcal{L}_3 := \mathcal{T}^{-1}\mathcal{L}_2\mathcal{T} = \Pi_S^\perp\big(\omega\cdot\partial_\vartheta + m_3\partial_{zzz} + d_1\partial_z + d_0\big)\Pi_S^\perp + \mathfrak{R}_3\,, \tag{8.65}$$

$$d_1 := (\mathcal{T}^{-1}c_1) + \omega\cdot\partial_\vartheta p\,, \qquad d_0 := \mathcal{T}^{-1}c_0\,, \qquad \mathfrak{R}_3 := \mathcal{T}^{-1}\mathfrak{R}_2\mathcal{T}. \tag{8.66}$$

We choose

$$m_1 := \frac{1}{(2\pi)^{\nu+1}} \int_{\mathbb{T}^{\nu+1}} c_1 \, d\vartheta\, dy\,, \quad p := (\omega\cdot\partial_\vartheta)^{-1}\Big(m_1 - \frac{1}{2\pi}\int_{\mathbb{T}} c_1\, dy\Big)\,, \tag{8.67}$$

so that $\frac{1}{2\pi}\int_{\mathbb{T}} d_1(\vartheta, z)\, dz = m_1$ for all $\vartheta \in \mathbb{T}^\nu$. Note that, by (8.59),

$$\int_{\mathbb{T}} c_1(\vartheta, y)\, dy = \int_{\mathbb{T}} c_{1,\geq 4}(\vartheta, y)\, dy\,, \quad \omega\cdot\partial_\vartheta p(\vartheta) = m_1 - \frac{1}{2\pi}\int_{\mathbb{T}} c_{1,\geq 4}(\vartheta, y)\, dy \tag{8.68}$$

because $\bar p_1$, $\bar p_2$, q_{c_1} have all zero space-average. Also note that $\mathfrak{R}_3$ has the form (7.7). Since $\mathcal{T}$ is symplectic, the operator $\mathcal{L}_3$ in (8.65) is Hamiltonian.

Remark 8.11. We require Hypothesis (S1) so that the function $q_{>2}$ has zero space average (see Lemma 7.5). If $q_{>2}$ did not have zero average, then p in (8.67) would have size $O(\varepsilon^3\gamma^{-1})$ (see (7.31)) and, since $\mathcal{T}^{-1} - I = O(\varepsilon^3\gamma^{-1})$, the function $\tilde d_0$ in (8.71) would satisfy $\tilde d_0 = O(\varepsilon^4\gamma^{-1})$. Therefore it would remain a term of order ∂_x^0 which is not perturbative for the reducibility scheme of Section 8.7.

We put in evidence the terms of size ε, ε^2 in d_0, d_1, $\mathfrak{R}_3$. Recalling (8.66), (8.59), we split

$$d_1 = -\varepsilon\bar p_1 - \varepsilon^2\bar p_2 + \tilde d_1\,, \quad d_0 = -\varepsilon(\bar p_1)_x - \varepsilon^2(\bar p_2)_x + \tilde d_0\,, \quad \mathfrak{R}_3 = -\varepsilon^2\Pi_S^\perp\partial_x\bar{\mathcal{R}}_2 + \widetilde{\mathcal{R}}_* \tag{8.69}$$

where $\bar{\mathcal{R}}_2$ is obtained replacing v_δ with $\bar v$ in $\mathcal{R}_2$ (see (7.29)), and

$$\tilde d_1 := \varepsilon(\bar p_1 - \mathcal{T}^{-1}\bar p_1) + \varepsilon^2(\bar p_2 - \mathcal{T}^{-1}\bar p_2) + \mathcal{T}^{-1}(q_{c_1} + c_{1,\geq 4}) + \omega\cdot\partial_\vartheta p, \tag{8.70}$$

$$\tilde d_0 := \varepsilon(\bar p_1 - \mathcal{T}^{-1}\bar p_1)_x + \varepsilon^2(\bar p_2 - \mathcal{T}^{-1}\bar p_2)_x + \mathcal{T}^{-1}(q_{c_0} + c_{0,\geq 4}), \tag{8.71}$$

$$\widetilde{\mathcal{R}}_* := \mathcal{T}^{-1}\mathcal{R}_*\mathcal{T} + \varepsilon^2\Pi_S^\perp\partial_x(\mathcal{R}_2 - \mathcal{T}^{-1}\mathcal{R}_2\mathcal{T}) + \varepsilon^2\Pi_S^\perp\partial_x(\bar{\mathcal{R}}_2 - \mathcal{R}_2), \tag{8.72}$$

and $\mathcal{R}_*$ is defined in (8.49). We have also used that $\mathcal{T}^{-1}$ commutes with ∂_x and with $\Pi_S^\perp$.

Remark 8.12. The space average $\frac{1}{2\pi}\int_{\mathbb{T}}\tilde d_1(\vartheta,z)\,dz = \frac{1}{2\pi}\int_{\mathbb{T}} d_1(\vartheta,z)\,dz = m_1$ for all $\vartheta\in\mathbb{T}^\nu$.

Lemma 8.13. *There is $\sigma := \sigma(\nu,\tau) > 0$ (possibly larger than in Lemma 8.7) such that*

$$|m_1|^{\mathrm{Lip}(\gamma)} \leq C\varepsilon^4, \quad |\partial_i m_1[\widehat\imath]| \leq C\varepsilon^{b+2}\|\widehat\imath\|_{s_0+\sigma} \tag{8.73}$$

$$\|p\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^4\gamma^{-1} + \varepsilon^{b+2}\gamma^{-1}\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \quad \|\partial_i p[\widehat\imath]\|_s \leq_s \varepsilon^{b+2}\gamma^{-1}\big(\|\widehat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat\imath\|_{s_0+\sigma}\big), \tag{8.74}$$

$$\|\tilde d_k\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^5\gamma^{-1} + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \quad \|\partial_i\tilde d_k[\widehat\imath]\|_s \leq_s \varepsilon\big(\|\widehat\imath\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat\imath\|_{s_0+\sigma}\big) \tag{8.75}$$

for $k = 0, 1$. Moreover the matrix s-decay norm (see (2.16))

$$|\widetilde{\mathcal{R}}_*|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^3 + \varepsilon^2\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \quad |\partial_i\widetilde{\mathcal{R}}_*[\widehat\imath]|_s \leq_s \varepsilon^2\|\widehat\imath\|_{s+\sigma} + \varepsilon^{2b-1}\|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat\imath\|_{s_0+\sigma}\,. \tag{8.76}$$

The transformations $\mathcal{T}$, $\mathcal{T}^{-1}$ satisfy (8.39), (8.40).

Proof. The estimates (8.73), (8.74) follow by (8.67), (8.59), (8.68), and the bounds for $c_{1,\geq 4}$, $c_{0,\geq 4}$ in Lemma 8.7. The estimates (8.75) follow similarly by (8.63), (8.68), (8.74). The estimates (8.76) follow because $\mathcal{T}^{-1}\mathcal{R}_*\mathcal{T}$ satisfies the bounds (7.41) like $\mathcal{R}_*$ does (use Lemma 7.3 and (8.74)) and $|\varepsilon^2\Pi_S^\perp\partial_x(\bar{\mathcal{R}}_2 - \mathcal{R}_2)|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^2\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}$. $\square$

It is sufficient to estimate $\widetilde{\mathcal{R}}_*$ (which has the form (7.7)) only in the s-decay norm (see (8.76)) because the next transformations will preserve it. Such norms are used in the reducibility scheme of Section 8.7.

8.4. Linear Birkhoff normal form. Step 1

Now we eliminate the terms of order ε and ε^2 of $\mathcal{L}_3$. This step is different from the reducibility steps that we shall perform in Section 8.7, because the diophantine constant $\gamma = o(\varepsilon^2)$ (see (5.4)) and so terms $O(\varepsilon)$, $O(\varepsilon^2)$ are not perturbative. This reduction is possible thanks to the special form of the terms $\varepsilon\mathcal{B}_1$, $\varepsilon^2\mathcal{B}_2$ defined in (8.77): the harmonics of $\varepsilon\mathcal{B}_1$, and $\varepsilon^2 T$ in (8.93), which correspond to a possible small divisor are naught, see Corollary 8.17, and Lemma 8.21. In this section we eliminate the term $\varepsilon\mathcal{B}_1$. In Section 8.5 we eliminate the terms of order ε^2.

Note that, since the previous transformations Φ, B, $\mathcal{T}$ are $O(\varepsilon^4\gamma^{-1})$-close to the identity, the terms of order ε and ε^2 in $\mathcal{L}_3$ are the same as in the original linearized operator.

We first collect all the terms of order ε and ε^2 in the operator $\mathcal{L}_3$ defined in (8.65). By (8.69), (7.29), (8.57) we have, renaming $\vartheta = \varphi$, $z = x$,

$$\mathcal{L}_3 = \Pi_S^\perp\big(\omega\cdot\partial_\varphi + m_3\partial_{xxx} + \varepsilon\mathcal{B}_1 + \varepsilon^2\mathcal{B}_2 + \tilde d_1\partial_x + \tilde d_0\big)\Pi_S^\perp + \widetilde{\mathcal{R}}_*$$

where $\tilde{d}_1$, $\tilde{d}_0$, $\widetilde{\mathcal{R}}_*$ are defined in (8.70)–(8.72) and (recall also (2.2))

$$\mathcal{B}_1 h := -6\partial_x(\bar{v}h), \quad \mathcal{B}_2 h := -6\partial_x\{\bar{v}\Pi_S[(\partial_x^{-1}\bar{v})\,\partial_x^{-1}h] + h\pi_0[(\partial_x^{-1}\bar{v})^2]\} + 6\pi_0\{(\partial_x^{-1}\bar{v})\Pi_S[\bar{v}h]\}. \tag{8.77}$$

Note that $\mathcal{B}_1$ and $\mathcal{B}_2$ are the linear Hamiltonian vector fields of $H_S^{\perp}$ generated, respectively, by the Hamiltonian $z \mapsto 3\int_{\mathbb{T}} vz^2$ in (3.6), and the fourth order Birkhoff Hamiltonian $\mathcal{H}_{4,2}$ in (3.7) at $v = \bar{v}$.

We transform $\mathcal{L}_3$ by a symplectic operator $\Phi_1 : H^s_{S^\perp}(\mathbb{T}^{\nu+1}) \to H^s_{S^\perp}(\mathbb{T}^{\nu+1})$ of the form

$$\Phi_1 := \exp(\varepsilon A_1) = I_{H_S^\perp} + \varepsilon A_1 + \varepsilon^2\frac{A_1^2}{2} + \varepsilon^3\widehat{A}_1, \quad \widehat{A}_1 := \sum_{k\ge 3}\frac{\varepsilon^{k-3}}{k!}A_1^k, \tag{8.78}$$

where $A_1(\varphi)h = \sum_{j,j'\in S^c}(A_1)_j^{j'}(\varphi)h_{j'}e^{\mathrm{i}jx}$ is a Hamiltonian vector field. The map Φ_1 is symplectic, because it is the time-1 flow of a Hamiltonian vector field. Therefore

$$\begin{aligned}\mathcal{L}_3\Phi_1 - \Phi_1\Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx})\Pi_S^\perp = \Pi_S^\perp(&\varepsilon\{\mathcal{D}_\omega A_1 + m_3[\partial_{xxx}, A_1] + \mathcal{B}_1\}\\ &+ \varepsilon^2\{\mathcal{B}_1A_1 + \mathcal{B}_2 + \frac{1}{2}m_3[\partial_{xxx}, A_1^2] + \frac{1}{2}(\mathcal{D}_\omega A_1^2)\}\\ &+ \tilde{d}_1\partial_x + R_3)\Pi_S^\perp\end{aligned} \tag{8.79}$$

where

$$\begin{aligned}R_3 := {}& \tilde{d}_1\partial_x(\Phi_1 - I) + \tilde{d}_0\Phi_1 + \widetilde{\mathcal{R}}_*\Phi_1 + \varepsilon^2\mathcal{B}_2(\Phi_1 - I)\\ &+ \varepsilon^3\big\{\mathcal{D}_\omega\widehat{A}_1 + m_3[\partial_{xxx}, \widehat{A}_1] + \frac{1}{2}\mathcal{B}_1A_1^2 + \varepsilon\mathcal{B}_1\widehat{A}_1\big\}.\end{aligned} \tag{8.80}$$

Remark 8.14. R_3 has no longer the form (7.7). However $R_3 = O(\partial_x^0)$ because $A_1 = O(\partial_x^{-1})$ (see Lemma 8.19), and therefore $\Phi_1 - I_{H_S^\perp} = O(\partial_x^{-1})$. Moreover the matrix decay norm of R_3 is $o(\varepsilon^2)$.

In order to eliminate the order ε from (8.79), we choose

$$(A_1)_j^{j'}(l) := \begin{cases} -\dfrac{(\mathcal{B}_1)_j^{j'}(l)}{\mathrm{i}(\omega\cdot l + m_3(j'^3 - j^3))} & \text{if } \bar{\omega}\cdot l + j'^3 - j^3 \neq 0, \\ 0 & \text{otherwise,}\end{cases} \qquad j, j' \in S^c,\ l \in \mathbb{Z}^\nu. \tag{8.81}$$

This definition is well posed. Indeed, by (8.77) and (8.57),

$$(\mathcal{B}_1)_j^{j'}(l) := \begin{cases} -6\mathrm{i}j\sqrt{\xi_{j-j'}} & \text{if } j - j' \in S,\ l = \ell(j - j') \\ 0 & \text{otherwise.}\end{cases} \tag{8.82}$$

In particular $(\mathcal{B}_1)_j^{j'}(l) = 0$ unless $|l| \le 1$. Thus, for $\bar{\omega}\cdot l + j'^3 - j^3 \neq 0$, the denominators in (8.81) satisfy

$$\begin{aligned}|\omega\cdot l + m_3(j'^3 - j^3)| &= |m_3(\bar{\omega}\cdot l + j'^3 - j^3) + (\omega - m_3\bar{\omega})\cdot l|\\ &\ge |m_3||\bar{\omega}\cdot l + j'^3 - j^3| - |\omega - m_3\bar{\omega}||l| \ge 1/2, \quad \forall |l| \le 1,\end{aligned} \tag{8.83}$$

for ε small, because the nonzero integer $|\bar{\omega}\cdot l + j'^3 - j^3| \ge 1$, (8.50), and $\omega = \bar{\omega} + O(\varepsilon^2)$.

A_1 defined in (8.81) is a Hamiltonian vector field like $\mathcal{B}_1$.

Remark 8.15. This is a general fact: the denominators $\delta_{l,j,k} := \mathrm{i}(\omega\cdot l + m_3(k^3 - j^3))$ satisfy $\overline{\delta_{l,j,k}} = \delta_{-l,k,j}$ and an operator $G(\varphi)$ is self-adjoint if and only if its matrix elements satisfy $\overline{G_j^k(l)} = G_k^j(-l)$, see [2]-Remark 4.5. In a more intrinsic way, we could solve the homological equation of this Birkhoff step directly for the Hamiltonian function whose flow generates Φ_1.

Lemma 8.16. *If $j, j' \in S^c$, $j - j' \in S$, $l = \ell(j - j')$, then $\bar{\omega}\cdot l + j'^3 - j^3 = 3jj'(j' - j) \neq 0$.*

Proof. We have $\bar\omega \cdot l = \bar\omega \cdot \ell(j-j') = (j-j')^3$ because $j-j' \in S$ (see (2.10) and (8.58)). Note that $j, j' \neq 0$ because $j, j' \in S^c$, and $j - j' \neq 0$ because $j - j' \in S$. $\square$

Corollary 8.17. *Let $j, j' \in S^c$. If $\bar\omega \cdot l + j'^3 - j^3 = 0$ then $(\mathcal{B}_1)_j^{j'}(l) = 0$.*

Proof. If $(\mathcal{B}_1)_j^{j'}(l) \neq 0$ then $j - j' \in S, l = \ell(j-j')$ by (8.82). Hence $\bar\omega \cdot l + j'^3 - j^3 \neq 0$ by Lemma 8.16. $\square$

By (8.81) and the previous corollary, the term of order ε in (8.79) is

$$\Pi_S^\perp \big(\mathcal{D}_\omega A_1 + m_3[\partial_{xxx}, A_1] + \mathcal{B}_1\big)\Pi_S^\perp = 0\,. \tag{8.84}$$

We now estimate the transformation A_1.

Lemma 8.18. (i) *For all $l \in \mathbb{Z}^\nu$, $j, j' \in S^c$,*

$$|(A_1)_j^{j'}(l)| \le C(|j|+|j'|)^{-1}\,, \quad |(A_1)_j^{j'}(l)|^{\mathrm{lip}} \le \varepsilon^{-2}(|j|+|j'|)^{-1}\,. \tag{8.85}$$

(ii) $(A_1)_j^{j'}(l) = 0$ *for all $l \in \mathbb{Z}^\nu$, $j, j' \in S^c$ such that $|j-j'| > C_S$, where $C_S := \max\{|j| : j \in S\}$.*

Proof. (i) We already noted that $(A_1)_j^{j'}(l) = 0$, $\forall |l| > 1$. Since $|\omega| \le |\bar\omega| + 1$, one has, for $|l| \le 1$, $j \neq j'$,

$$|\omega \cdot l + m_3(j'^3 - j^3)| \ge |m_3|\,|j'^3 - j^3| - |\omega \cdot l| \ge \frac{1}{4}(j'^2 + j^2) - |\omega| \ge \frac{1}{8}(j'^2+j^2)\,, \quad \forall (j'^2+j^2) \ge C,$$

for some constant $C > 0$. Moreover, recalling that also (8.83) holds, we deduce that for $j \neq j'$,

$$(A_1)_j^{j'}(l) \neq 0 \quad \Rightarrow \quad |\omega \cdot l + m_3(j'^3 - j^3)| \ge c(|j|+|j'|)^2\,. \tag{8.86}$$

On the other hand, if $j = j'$, $j \in S^c$, the matrix $(A_1)_j^j(l) = 0$, $\forall l \in \mathbb{Z}^\nu$, because $(\mathcal{B}_1)_j^j(l) = 0$ by (8.82) (recall that $0 \notin S$). Hence (8.86) holds for all j, j'. By (8.81), (8.86), (8.82) we deduce the first bound in (8.85). The Lipschitz bound follows similarly (use also $|j-j'| \le C_S$). (ii) follows by (8.81)–(8.82). $\square$

The previous lemma means that $A = O(|\partial_x|^{-1})$. More precisely we deduce that

Lemma 8.19. $|A_1\partial_x|_s^{\mathrm{Lip}(\gamma)} + |\partial_x A_1|_s^{\mathrm{Lip}(\gamma)} \le C(s)$.

Proof. Recalling the definition of the (space–time) matrix norm in (2.23), since $(A_1)_{j_1}^{j_2}(l) = 0$ outside the set of indices $|l| \le 1, |j_1 - j_2| \le C_S$, we have

$$|\partial_x A_1|_s^2 = \sum_{|l|\le 1,\, |j| \le C_S} \Big(\sup_{j_1 - j_2 = j} |j_1|\,|(A_1)_{j_1}^{j_2}(l)| \Big)^2 \langle l, j\rangle^{2s} \le C(s)$$

by Lemma 8.18. The estimates for $|A_1\partial_x|_s$ and the Lipschitz bounds follow similarly. $\square$

It follows that the symplectic map Φ_1 in (8.78) is invertible for ε small, with inverse

$$\Phi_1^{-1} = \exp(-\varepsilon A_1) = I_{H_S^\perp} + \varepsilon \check A_1\,, \ \ \check A_1 := \sum_{n\ge1} \frac{\varepsilon^{n-1}}{n!}(-A_1)^n\,, \ \ |\check A_1 \partial_x|_s^{\mathrm{Lip}(\gamma)} + |\partial_x \check A_1|_s^{\mathrm{Lip}(\gamma)} \le C(s)\,. \tag{8.87}$$

Since A_1 solves the homological equation (8.84), the ε-term in (8.79) is zero, and, with a straightforward calculation, the ε^2-term simplifies to $\mathcal{B}_2 + \frac{1}{2}[\mathcal{B}_1, A_1]$. We obtain the Hamiltonian operator

$$\mathcal{L}_4 := \Phi_1^{-1}\mathcal{L}_3\Phi_1 = \Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx} + \tilde d_1 \partial_x + \varepsilon^2\{\mathcal{B}_2 + \tfrac{1}{2}[\mathcal{B}_1, A_1]\} + \tilde R_4)\Pi_S^\perp \tag{8.88}$$

$$\tilde R_4 := (\Phi_1^{-1} - I)\Pi_S^\perp[\varepsilon^2(\mathcal{B}_2 + \tfrac{1}{2}[\mathcal{B}_1, A_1]) + \tilde d_1\partial_x] + \Phi_1^{-1}\Pi_S^\perp R_3\,. \tag{8.89}$$

We split A_1 defined in (8.81), (8.82) into $A_1 = \bar{A}_1 + \widetilde{A}_1$ where, for all $j, j' \in S^c$, $l \in \mathbb{Z}^\nu$,

$$(\bar{A}_1)_j^{j'}(l) := \frac{6j\sqrt{\xi_{j-j'}}}{\bar{\omega}\cdot l + j'^3 - j^3} \qquad \text{if } \bar{\omega}\cdot l + j'^3 - j^3 \neq 0, \;\; j - j' \in S, \;\; l = \ell(j-j'), \tag{8.90}$$

and $(\bar{A}_1)_j^{j'}(l) := 0$ otherwise. By Lemma 8.16, for all $j, j' \in S^c$, $l \in \mathbb{Z}^\nu$, $(\bar{A}_1)_j^{j'}(l) = \frac{2\sqrt{\xi_{j-j'}}}{j'(j'-j)}$ if $j - j' \in S$, $l = \ell(j-j')$, and $(\bar{A}_1)_j^{j'}(l) = 0$ otherwise, namely (recall the definition of $\bar{v}$ in (8.57))

$$\bar{A}_1 h = 2\Pi_S^\perp[(\partial_x^{-1}\bar{v})(\partial_x^{-1}h)], \quad \forall h \in H^s_{S^\perp}(\mathbb{T}^{\nu+1}). \tag{8.91}$$

The difference is

$$(\widetilde{A}_1)_j^{j'}(l) = (A_1 - \bar{A}_1)_j^{j'}(l) = -\frac{6j\sqrt{\xi_{j-j'}}\{(\omega-\bar{\omega})\cdot l + (m_3 - 1)(j'^3 - j^3)\}}{\big(\omega\cdot l + m_3(j'^3 - j^3)\big)\big(\bar{\omega}\cdot l + j'^3 - j^3\big)} \tag{8.92}$$

for $j, j' \in S^c$, $j - j' \in S$, $l = \ell(j-j')$, and $(\widetilde{A}_1)_j^{j'}(l) = 0$ otherwise. Then, by (8.88),

$$\mathcal{L}_4 = \Pi_S^\perp\big(\mathcal{D}_\omega + m_3\partial_{xxx} + \tilde{d}_1\partial_x + \varepsilon^2 T + R_4\big)\Pi_S^\perp, \tag{8.93}$$

where

$$T := \mathcal{B}_2 + \frac{1}{2}[\mathcal{B}_1, \bar{A}_1], \qquad R_4 := \frac{\varepsilon^2}{2}[\mathcal{B}_1, \widetilde{A}_1] + \tilde{R}_4. \tag{8.94}$$

The operator T is Hamiltonian like $\mathcal{B}_2$, $\mathcal{B}_1$, $\bar{A}_1$ (the commutator of two Hamiltonian vector fields is Hamiltonian).

Lemma 8.20. *There is $\sigma = \sigma(\nu, \tau) > 0$ (possibly larger than in Lemma 8.13) such that*

$$|R_4|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^5\gamma^{-1} + \varepsilon\|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \quad |\partial_i R_4[\widehat{\imath}]|_s \leq_s \varepsilon\big(\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}\big). \tag{8.95}$$

Proof. We first estimate $[\mathcal{B}_1, \widetilde{A}_1] = (\mathcal{B}_1\partial_x^{-1})(\partial_x\widetilde{A}_1) - (\widetilde{A}_1\partial_x)(\partial_x^{-1}\mathcal{B}_1)$. By (8.92), $|\omega - \bar{\omega}| \leq C\varepsilon^2$ (as $\omega \in \Omega_\varepsilon$ in (5.2)) and (8.50), arguing as in Lemmata 8.18, 8.19, we deduce that $|\widetilde{A}_1\partial_x|_s^{\mathrm{Lip}(\gamma)} + |\partial_x\widetilde{A}_1|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^2$. By (8.77) the norm $|\mathcal{B}_1\partial_x^{-1}|_s^{\mathrm{Lip}(\gamma)} + |\partial_x^{-1}\mathcal{B}_1|^{\mathrm{Lip}(\gamma)} \leq C(s)$. Hence $\varepsilon^2|[\mathcal{B}_1, \widetilde{A}_1]|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^4$. Finally (8.94), (8.89), (8.87), (8.80), (8.75), (8.76), and the interpolation estimate (2.20) imply (8.95). □

8.5. Linear Birkhoff normal form. Step 2

The goal of this section is to remove the term $\varepsilon^2 T$ from the operator $\mathcal{L}_4$ defined in (8.93). We conjugate the Hamiltonian operator $\mathcal{L}_4$ via a symplectic map

$$\Phi_2 := \exp(\varepsilon^2 A_2) = I_{H_S^\perp} + \varepsilon^2 A_2 + \varepsilon^4\widehat{A}_2, \quad \widehat{A}_2 := \sum_{k\geq 2}\frac{\varepsilon^{2(k-2)}}{k!}A_2^k \tag{8.96}$$

where $A_2(\varphi) = \sum_{j,j'\in S^c}(A_2)_j^{j'}(\varphi)h_{j'}e^{\mathrm{i}jx}$ is a Hamiltonian vector field. We compute

$$\mathcal{L}_4\Phi_2 - \Phi_2\Pi_S^\perp\big(\mathcal{D}_\omega + m_3\partial_{xxx}\big)\Pi_S^\perp = \Pi_S^\perp(\varepsilon^2\{\mathcal{D}_\omega A_2 + m_3[\partial_{xxx}, A_2] + T\} + \tilde{d}_1\partial_x + \tilde{R}_5)\Pi_S^\perp, \tag{8.97}$$

$$\tilde{R}_5 := \Pi_S^\perp\{\varepsilon^4((\mathcal{D}_\omega\widehat{A}_2) + m_3[\partial_{xxx}, \widehat{A}_2]) + (\tilde{d}_1\partial_x + \varepsilon^2 T)(\Phi_2 - I) + R_4\Phi_2\}\Pi_S^\perp. \tag{8.98}$$

We define

$$(A_2)_j^{j'}(l) := -\frac{T_j^{j'}(l)}{\mathrm{i}(\omega\cdot l + m_3(j'^3 - j^3))} \quad \text{if } \bar{\omega}\cdot l + j'^3 - j^3 \neq 0; \qquad (A_2)_j^{j'}(l) := 0 \quad \text{otherwise.} \tag{8.99}$$

This definition is well posed. Indeed, by (8.94), (8.82), (8.90), (8.77), the matrix entries $T_j^{j'}(l) = 0$ for all $|j - j'| > 2C_S$, $l \in \mathbb{Z}^\nu$, where $C_S := \max\{|j|, j \in S\}$. Also $T_j^{j'}(l) = 0$ for all $j, j' \in S^c$, $|l| > 2$ (see also (8.100), (8.103), (8.104) below). Thus, arguing like in (8.83), if $\bar\omega \cdot l + j'^3 - j^3 \neq 0$, then $|\omega \cdot l + m_3(j'^3 - j^3)| \geq 1/2$. The operator A_2 is a Hamiltonian vector field because T is Hamiltonian and by Remark 8.15.

Now we prove that the Birkhoff map Φ_2 removes completely the term $\varepsilon^2 T$.

Lemma 8.21. *Let $j, j' \in S^c$. If $\bar\omega \cdot l + j'^3 - j^3 = 0$, then $T_j^{j'}(l) = 0$.*

Proof. By (8.77), (8.91) we get $\mathcal{B}_1 \bar A_1 h = -12\partial_x\{\bar v \Pi_S^\perp[(\partial_x^{-1}\bar v)(\partial_x^{-1}h)]\}$, $\bar A_1 \mathcal{B}_1 h = -12\Pi_S^\perp[(\partial_x^{-1}\bar v)\Pi_S^\perp(\bar v h)]$ for all $h \in H^s_{S^\perp}$, whence, recalling (8.57), for all $j, j' \in S^c$, $l \in \mathbb{Z}^\nu$,

$$([\mathcal{B}_1, \bar A_1])_j^{j'}(l) = 12\mathrm{i} \sum_{\substack{j_1, j_2 \in S,\ j_1 + j_2 = j - j' \\ j' + j_2 \in S^c,\ \ell(j_1) + \ell(j_2) = l}} \frac{j j_1 - j' j_2}{j' j_1 j_2} \sqrt{\xi_{j_1}\xi_{j_2}}. \tag{8.100}$$

If $([\mathcal{B}_1, \bar A_1])_j^{j'}(l) \neq 0$ there are $j_1, j_2 \in S$ such that $j_1 + j_2 = j - j'$, $j' + j_2 \in S^c$, $\ell(j_1) + \ell(j_2) = l$. Then

$$\bar\omega \cdot l + j'^3 - j^3 = \bar\omega \cdot \ell(j_1) + \bar\omega \cdot \ell(j_2) + j'^3 - j^3 \overset{(8.58)}{=} j_1^3 + j_2^3 + j'^3 - j^3. \tag{8.101}$$

Thus, if $\bar\omega \cdot l + j'^3 - j^3 = 0$, Lemma 3.3 implies $(j_1 + j_2)(j_1 + j')(j_2 + j') = 0$. Now $j_1 + j'$, $j_2 + j' \neq 0$ because $j_1, j_2 \in S$, $j' \in S^c$ and S is symmetric. Hence $j_1 + j_2 = 0$, which implies $j = j'$ and $l = 0$ (the map ℓ in (8.58) is odd). In conclusion, if $\bar\omega \cdot l + j'^3 - j^3 = 0$, the only nonzero matrix entry $([\mathcal{B}_1, \bar A_1])_j^{j'}(l)$ is

$$([\mathcal{B}_1, \bar A_1])_j^{j}(0) \overset{(8.100)}{=} 24\mathrm{i} \sum_{j_2 \in S,\ j_2 + j \in S^c} \xi_{j_2} j_2^{-1}. \tag{8.102}$$

Now we consider $\mathcal{B}_2$ in (8.77). Split $\mathcal{B}_2 = B_1 + B_2 + B_3$, where $B_1 h := -6\partial_x\{\bar v \Pi_S[(\partial_x^{-1}\bar v)\partial_x^{-1}h]\}$, $B_2 h := -6\partial_x\{h\pi_0[(\partial_x^{-1}\bar v)^2]\}$, $B_3 h := 6\pi_0\{\Pi_S(\bar v h)\partial_x^{-1}\bar v\}$. Their Fourier matrix representation is

$$(B_1)_j^{j'}(l) = 6\mathrm{i} j \sum_{\substack{j_1, j_2 \in S,\ j_1 + j' \in S \\ j_1 + j_2 = j - j',\ \ell(j_1) + \ell(j_2) = l}} \frac{\sqrt{\xi_{j_1}\xi_{j_2}}}{j_1 j'},$$

$$(B_2)_j^{j'}(l) = 6\mathrm{i} j \sum_{\substack{j_1, j_2 \in S,\ j_1 + j_2 \neq 0 \\ j_1 + j_2 = j - j',\ \ell(j_1) + \ell(j_2) = l}} \frac{\sqrt{\xi_{j_1}\xi_{j_2}}}{j_1 j_2}, \tag{8.103}$$

$$(B_3)_j^{j'}(l) = 6 \sum_{\substack{j_1, j_2 \in S,\ j_1 + j' \in S \\ j_1 + j_2 = j - j',\ \ell(j_1) + \ell(j_2) = l}} \frac{\sqrt{\xi_{j_1}\xi_{j_2}}}{\mathrm{i} j_2}, \qquad j, j' \in S^c,\ l \in \mathbb{Z}^\nu. \tag{8.104}$$

We study the terms B_1, B_2, B_3 separately. If $(B_1)_j^{j'}(l) \neq 0$, there are $j_1, j_2 \in S$ such that $j_1 + j_2 = j - j'$, $j_1 + j' \in S$, $l = \ell(j_1) + \ell(j_2)$ and (8.101) holds. Thus, if $\bar\omega \cdot l + j'^3 - j^3 = 0$, Lemma 3.3 implies $(j_1 + j_2)(j_1 + j')(j_2 + j') = 0$, and, since $j' \in S^c$ and S is symmetric, the only possibility is $j_1 + j_2 = 0$. Hence $j = j'$, $l = 0$. In conclusion, if $\bar\omega \cdot l + j'^3 - j^3 = 0$, the only nonzero matrix element $(B_1)_j^{j'}(l)$ is

$$(B_1)_j^j(0) = 6\mathrm{i} \sum_{j_1 \in S,\ j_1 + j \in S} \xi_{j_1} j_1^{-1}. \tag{8.105}$$

By the same arguments, if $(B_2)_j^{j'}(l) \neq 0$ and $\bar\omega \cdot l + j'^3 - j^3 = 0$ we find $(j_1 + j_2)(j_1 + j')(j_2 + j') = 0$, which is impossible because also $j_1 + j_2 \neq 0$. Finally, arguing as for B_1, if $\bar\omega \cdot l + j'^3 - j^3 = 0$, then the only nonzero matrix element $(B_3)_j^{j'}(l)$ is

$$(B_3)_j^j(0) = 6\mathrm{i} \sum_{j_1 \in S,\ j_1 + j \in S} \xi_{j_1} j_1^{-1}. \tag{8.106}$$

From (8.102), (8.105), (8.106) we deduce that, if $\bar{\omega} \cdot l + j'^3 - j^3 = 0$, then the only nonzero elements $(\frac{1}{2}[\mathcal{B}_1, \bar{A}_1] + B_1 + B_3)_j^{j'}(l)$ must be for $(l, j, j') = (0, j, j)$. In this case, we get

$$\frac{1}{2}([\mathcal{B}_1, \bar{A}_1])_j^j(0) + (B_1)_j^j(0) + (B_3)_j^j(0) = 12\mathrm{i} \sum_{\substack{j_1 \in S \\ j_1 + j \in S^c}} \frac{\xi_{j_1}}{j_1} + 12\mathrm{i} \sum_{\substack{j_1 \in S \\ j_1 + j \in S}} \frac{\xi_{j_1}}{j_1} = 12\mathrm{i} \sum_{j_1 \in S} \frac{\xi_{j_1}}{j_1} = 0 \tag{8.107}$$

because the case $j_1 + j = 0$ is impossible ($j_1 \in S$, $j' \in S^c$ and S is symmetric), and the function $S \ni j_1 \to \xi_{j_1}/j_1 \in \mathbb{R}$ is odd. The lemma follows by (8.94), (8.107). □

The choice of A_2 in (8.99) and Lemma 8.21 imply that

$$\Pi_S^\perp\big(\mathcal{D}_\omega A_2 + m_3[\partial_{xxx}, A_2] + T\big)\Pi_S^\perp = 0\,. \tag{8.108}$$

Lemma 8.22. $|\partial_x A_2|_s^{\mathrm{Lip}(\gamma)} + |A_2 \partial_x|_s^{\mathrm{Lip}(\gamma)} \le C(s)$.

Proof. First we prove that the diagonal elements $T_j^j(l) = 0$ for all $l \in \mathbb{Z}^\nu$. For $l = 0$, we have already proved that $T_j^j(0) = 0$ (apply Lemma 8.21 with $j = j'$, $l = 0$). Moreover, in each term $[\mathcal{B}_1, \bar{A}_1]$, B_1, B_2, B_3 (see (8.100), (8.103), (8.104)) the sum is over $j_1 + j_2 = j - j'$, $l = \ell(j_1) + \ell(j_2)$. If $j = j'$, then $j_1 + j_2 = 0$, and $l = 0$. Thus $T_j^j(l) = T_j^j(0) = 0$. For the off-diagonal terms $j \neq j'$ we argue as in Lemmata 8.18, 8.19, using that all the denominators $|\omega \cdot l + m_3(j'^3 - j^3)| \ge c(|j| + |j'|)^2$. □

For ε small, the map Φ_2 in (8.96) is invertible and $\Phi_2 = \exp(-\varepsilon^2 A_2)$. Therefore (8.97), (8.108) imply

$$\mathcal{L}_5 := \Phi_2^{-1} \mathcal{L}_4 \Phi_2 = \Pi_S^\perp(\mathcal{D}_\omega + m_3 \partial_{xxx} + \tilde{d}_1 \partial_x + R_5)\Pi_S^\perp\,, \tag{8.109}$$

$$R_5 := (\Phi_2^{-1} - I)\Pi_S^\perp \tilde{d}_1 \partial_x + \Phi_2^{-1} \Pi_S^\perp \tilde{R}_5\,. \tag{8.110}$$

Since A_2 is a Hamiltonian vector field, the map Φ_2 is symplectic and so $\mathcal{L}_5$ is Hamiltonian.

Lemma 8.23. *R_5 satisfies the same estimates (8.95) as R_4 (with a possibly larger σ).*

Proof. Use (8.110), Lemma 8.22, (8.75), (8.98), (8.95) and the interpolation inequalities (2.18), (2.20). □

8.6. Descent method

The goal of this section is to transform $\mathcal{L}_5$ in (8.109) so that the coefficient of ∂_x becomes constant. We conjugate $\mathcal{L}_5$ via a symplectic map of the form

$$\mathcal{S} := \exp(\Pi_S^\perp(w\partial_x^{-1}))\Pi_S^\perp = \Pi_S^\perp\big(I + w\partial_x^{-1}\big)\Pi_S^\perp + \widehat{\mathcal{S}}, \quad \widehat{\mathcal{S}} := \sum_{k \ge 2} \frac{1}{k!}[\Pi_S^\perp(w\partial_x^{-1})]^k \Pi_S^\perp\,, \tag{8.111}$$

where $w : \mathbb{T}^{\nu+1} \to \mathbb{R}$ is a function. Note that $\Pi_S^\perp(w\partial_x^{-1})\Pi_S^\perp$ is the Hamiltonian vector field generated by $-\frac{1}{2}\int_{\mathbb{T}} w(\partial_x^{-1}h)^2\,dx$, $h \in H_S^\perp$. Recalling (2.2), we calculate

$$\begin{aligned}
&\mathcal{L}_5\mathcal{S} - \mathcal{S}\Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx} + m_1\partial_x)\Pi_S^\perp = \Pi_S^\perp(3m_3 w_x + \tilde{d}_1 - m_1)\partial_x \Pi_S^\perp + \tilde{R}_6\,, \\
&\tilde{R}_6 := \Pi_S^\perp\{(3m_3 w_{xx} + \tilde{d}_1 \Pi_S^\perp w - m_1 w)\pi_0 + ((\mathcal{D}_\omega w) + m_3 w_{xxx} + \tilde{d}_1 \Pi_S^\perp w_x)\partial_x^{-1} + (\mathcal{D}_\omega \widehat{\mathcal{S}}) \\
&\qquad + m_3[\partial_{xxx}, \widehat{\mathcal{S}}] + \tilde{d}_1 \partial_x \widehat{\mathcal{S}} - m_1 \widehat{\mathcal{S}}\partial_x + R_5\mathcal{S}\}\Pi_S^\perp
\end{aligned} \tag{8.112}$$

where $\tilde{R}_6$ collects all the terms of order at most ∂_x^0. By Remark 8.12, we solve $3m_3 w_x + \tilde{d}_1 - m_1 = 0$ by choosing $w := -(3m_3)^{-1}\partial_x^{-1}(\tilde{d}_1 - m_1)$. For ε small, the operator $\mathcal{S}$ is invertible and, by (8.112),

$$\mathcal{L}_6 := \mathcal{S}^{-1}\mathcal{L}_5\mathcal{S} = \Pi_S^\perp(\mathcal{D}_\omega + m_3\partial_{xxx} + m_1\partial_x)\Pi_S^\perp + R_6\,, \qquad R_6 := \mathcal{S}^{-1}\tilde{R}_6\,. \tag{8.113}$$

Since $\mathcal{S}$ is symplectic, $\mathcal{L}_6$ is Hamiltonian (recall Definition 2.2).

Lemma 8.24. *There is $\sigma = \sigma(\nu, \tau) > 0$ (possibly larger than in Lemma 8.23) such that*

$$|\mathcal{S}^{\pm 1} - I|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^5 \gamma^{-1} + \varepsilon \|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}, \quad |\partial_i \mathcal{S}^{\pm 1}[\widehat{\imath}]|_s \leq_s \varepsilon(\|\widehat{\imath}\|_{s+\sigma} + \|\mathfrak{I}_\delta\|_{s+\sigma}\|\widehat{\imath}\|_{s_0+\sigma}).$$

The remainder R_6 satisfies the same estimates (8.95) as R_4.

Proof. By (8.75), (8.73), (8.50), $\|w\|_s^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^5 \gamma^{-1} + \varepsilon \|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}$, and the lemma follows by (8.111). Since $\widehat{\mathcal{S}} = O(\partial_x^{-2})$ the commutator $[\partial_{xxx}, \widehat{\mathcal{S}}] = O(\partial_x^0)$ and $|[\partial_{xxx}, \widehat{\mathcal{S}}]|_s^{\mathrm{Lip}(\gamma)} \leq_s \|w\|_{s_0+3}^{\mathrm{Lip}(\gamma)} \|w\|_{s+3}^{\mathrm{Lip}(\gamma)}$. □

8.7. KAM reducibility and inversion of $\mathcal{L}_\omega$

The coefficients m_3, m_1 of the operator $\mathcal{L}_6$ in (8.113) are constants, and the remainder R_6 is a bounded operator of order ∂_x^0 with small matrix decay norm, see (8.116). Then we can diagonalize $\mathcal{L}_6$ by applying the iterative KAM reducibility Theorem 4.2 in [2] along the sequence of scales

$$N_n := N_0^{\chi^n}, \quad n = 0, 1, 2, \ldots, \quad \chi := 3/2, \quad N_0 > 0. \tag{8.114}$$

In Section 9, the initial N_0 will (slightly) increase to infinity as $\varepsilon \to 0$, see (9.5). The required smallness condition (see (4.14) in [2]) is (written in the present notations)

$$N_0^{C_0} |R_6|_{s_0+\beta}^{\mathrm{Lip}(\gamma)} \gamma^{-1} \leq 1 \tag{8.115}$$

where $\beta := 7\tau + 6$ (see (4.1) in [2]), τ is the diophantine exponent in (5.4) and (8.120), and the constant $C_0 := C_0(\tau, \nu) > 0$ is fixed in Theorem 4.2 in [2]. By Lemma 8.24, the remainder R_6 satisfies the bound (8.95), and using (7.8) we get (recall (5.10))

$$|R_6|_{s_0+\beta}^{\mathrm{Lip}(\gamma)} \leq C\varepsilon^{7-2b}\gamma^{-1} = C\varepsilon^{3-2a}, \qquad |R_6|_{s_0+\beta}^{\mathrm{Lip}(\gamma)} \gamma^{-1} \leq C\varepsilon^{1-3a}. \tag{8.116}$$

We use that μ in (7.8) is assumed to satisfy $\mu \geq \sigma + \beta$ where $\sigma := \sigma(\tau, \nu)$ is given in Lemma 8.24.

Theorem 8.25 *(Reducibility). Assume that $\omega \mapsto i_\delta(\omega)$ is a Lipschitz function defined on some subset $\Omega_o \subset \Omega_\varepsilon$ (recall (5.2)), satisfying (7.8) with $\mu \geq \sigma + \beta$ where $\sigma := \sigma(\tau, \nu)$ is given in Lemma 8.24 and $\beta := 7\tau + 6$. Then there exists $\delta_0 \in (0, 1)$ such that, if*

$$N_0^{C_0} \varepsilon^{7-2b} \gamma^{-2} = N_0^{C_0} \varepsilon^{1-3a} \leq \delta_0, \quad \gamma := \varepsilon^{2+a}, \quad a \in (0, 1/6), \tag{8.117}$$

then:

(*i*) **(Eigenvalues)**. *For all $\omega \in \Omega_\varepsilon$ there exists a sequence*

$$\mu_j^\infty(\omega) := \mu_j^\infty(\omega, i_\delta(\omega)) := \mathrm{i}\big(-\tilde{m}_3(\omega) j^3 + \tilde{m}_1(\omega) j\big) + r_j^\infty(\omega), \quad j \in S^c, \tag{8.118}$$

where $\tilde{m}_3$, $\tilde{m}_1$ coincide with the coefficients m_3, m_1 of $\mathcal{L}_6$ in (8.113) for all $\omega \in \Omega_o$, and

$$|\tilde{m}_3 - 1|^{\mathrm{Lip}(\gamma)} + |\tilde{m}_1|^{\mathrm{Lip}(\gamma)} \leq C\varepsilon^4, \quad |r_j^\infty|^{\mathrm{Lip}(\gamma)} \leq C\varepsilon^{3-2a}, \quad \forall j \in S^c, \tag{8.119}$$

for some $C > 0$. All the eigenvalues μ_j^∞ are purely imaginary. We define, for convenience, $\mu_0^\infty(\omega) := 0$.

(*ii*) **(Conjugacy)**. *For all ω in the set*

$$\Omega_\infty^{2\gamma} := \Omega_\infty^{2\gamma}(i_\delta) := \left\{ \omega \in \Omega_o : |\mathrm{i}\omega \cdot l + \mu_j^\infty(\omega) - \mu_k^\infty(\omega)| \geq \frac{2\gamma |j^3 - k^3|}{\langle l \rangle^\tau}, \ \forall l \in \mathbb{Z}^\nu, \ j, k \in S^c \cup \{0\} \right\} \tag{8.120}$$

there is a real, bounded, invertible linear operator $\Phi_\infty(\omega) : H_{S^\perp}^s(\mathbb{T}^{\nu+1}) \to H_{S^\perp}^s(\mathbb{T}^{\nu+1})$, with bounded inverse $\Phi_\infty^{-1}(\omega)$, that conjugates $\mathcal{L}_6$ in (8.113) to constant coefficients, namely

$$\mathcal{L}_\infty(\omega) := \Phi_\infty^{-1}(\omega) \circ \mathcal{L}_6(\omega) \circ \Phi_\infty(\omega) = \omega \cdot \partial_\varphi + \mathcal{D}_\infty(\omega), \quad \mathcal{D}_\infty(\omega) := \mathrm{diag}_{j \in S^c}\{\mu_j^\infty(\omega)\}. \tag{8.121}$$

The transformations Φ_∞, Φ_∞^{-1} are close to the identity in matrix decay norm, with

$$|\Phi_\infty - I|_{s,\Omega_\infty^{2\gamma}}^{\mathrm{Lip}(\gamma)} + |\Phi_\infty^{-1} - I|_{s,\Omega_\infty^{2\gamma}}^{\mathrm{Lip}(\gamma)} \leq_s \varepsilon^5 \gamma^{-2} + \varepsilon \gamma^{-1} \|\mathfrak{I}_\delta\|_{s+\sigma}^{\mathrm{Lip}(\gamma)}. \tag{8.122}$$

Moreover Φ_∞, Φ_∞^{-1} are symplectic, and $\mathcal{L}_\infty$ is a Hamiltonian operator.

Proof. The proof is the same as the one of Theorem 4.1 in [2], which is based on Theorem 4.2, Corollaries 4.1, 4.2 and Lemmata 4.1, 4.2 of [2]. A difference is that here $\omega \in \mathbb{R}^\nu$, while in [2] the parameter $\lambda \in \mathbb{R}$ is one-dimensional. The proof is the same because Kirszbraun's Theorem on Lipschitz extension of functions also holds in $\mathbb{R}^\nu$ (see, e.g., Lemma A.2 in [27]). The bound (8.122) follows by Corollary 4.1 of [2] and the estimate of R_6 in Lemma 8.24. We also use the estimates (8.50), (8.73) for $\partial_i m_3$, $\partial_i m_1$ which correspond to (3.64) in [2]. Another difference is that here the sites $j \in S^c \subset \mathbb{Z} \setminus \{0\}$ unlike in [2] where $j \in \mathbb{Z}$. We have defined $\mu_0^\infty := 0$ so that also the first Melnikov conditions (8.123) are included in the definition of $\Omega_\infty^{2\gamma}$. □

Remark 8.26. Theorem 4.2 in [2] also provides the Lipschitz dependence of the (approximate) eigenvalues μ_j^n with respect to the unknown $i_0(\varphi)$, which is used for the measure estimate Lemma 9.3.

All the parameters $\omega \in \Omega_\infty^{2\gamma}$ satisfy (specialize (8.120) for $k = 0$)

$$|\mathrm{i}\omega \cdot l + \mu_j^\infty(\omega)| \geq 2\gamma |j|^3 \langle l \rangle^{-\tau}, \quad \forall l \in \mathbb{Z}^\nu, \ j \in S^c, \tag{8.123}$$

and the diagonal operator $\mathcal{L}_\infty$ is invertible.

In the following theorem we finally verify the inversion assumption (6.33) for $\mathcal{L}_\omega$.

Theorem 8.27 *(Inversion of $\mathcal{L}_\omega$). Assume the hypotheses of Theorem 8.25 and (8.117). Then there exists $\sigma_1 := \sigma_1(\tau, \nu) > 0$ such that, $\forall \omega \in \Omega_\infty^{2\gamma}(i_\delta)$ (see (8.120)), for any function $g \in H_{S^\perp}^{s+\sigma_1}(\mathbb{T}^{\nu+1})$ the equation $\mathcal{L}_\omega h = g$ has a solution $h = \mathcal{L}_\omega^{-1} g \in H_{S^\perp}^s(\mathbb{T}^{\nu+1})$, satisfying*

$$\|\mathcal{L}_\omega^{-1} g\|_s^{\mathrm{Lip}(\gamma)} \leq_s \gamma^{-1} \big(\|g\|_{s+\sigma_1}^{\mathrm{Lip}(\gamma)} + \varepsilon\gamma^{-1} \|\mathfrak{I}_0\|_{s+\sigma_1}^{\mathrm{Lip}(\gamma)} \|g\|_{s_0}^{\mathrm{Lip}(\gamma)} \big). \tag{8.124}$$

Proof. Collecting Theorem 8.25 with the results of Sections 8.1–8.6, we have obtained the (semi)-conjugation of the operator $\mathcal{L}_\omega$ (defined in (7.34)) to $\mathcal{L}_\infty$ (defined in (8.121)), namely

$$\mathcal{L}_\omega = \mathcal{M}_1 \mathcal{L}_\infty \mathcal{M}_2^{-1}, \qquad \mathcal{M}_1 := \Phi B \rho \mathcal{T} \Phi_1 \Phi_2 \mathcal{S} \Phi_\infty, \quad \mathcal{M}_2 := \Phi B \mathcal{T} \Phi_1 \Phi_2 \mathcal{S} \Phi_\infty, \tag{8.125}$$

where ρ means the multiplication operator by the function ρ defined in (8.41). By (8.123) and Lemma 4.2 of [2] we deduce that $\|\mathcal{L}_\infty^{-1} g\|_s^{\mathrm{Lip}(\gamma)} \leq_s \gamma^{-1} \|g\|_{s+2\tau+1}^{\mathrm{Lip}(\gamma)}$. In order to estimate $\mathcal{M}_2$, $\mathcal{M}_1^{-1}$, we recall that the composition of tame maps is tame, see Lemma 6.5 in [2]. Now, Φ, Φ^{-1} are estimated in Lemma 8.5, B, B^{-1} and ρ in Lemma 8.7, $\mathcal{T}$, $\mathcal{T}^{-1}$ in Lemma 8.13. The decay norms $|\Phi_1|_s^{\mathrm{Lip}(\gamma)}$, $|\Phi_1^{-1}|_s^{\mathrm{Lip}(\gamma)}$, $|\Phi_2|_s^{\mathrm{Lip}(\gamma)}$, $|\Phi_2^{-1}|_s^{\mathrm{Lip}(\gamma)} \leq C(s)$ by Lemmata 8.19, 8.22. The decay norm of $\mathcal{S}$, $\mathcal{S}^{-1}$ is estimated in Lemma 8.24, and Φ_∞, Φ_∞^{-1} in (8.122). The decay norm controls the Sobolev norm by (2.21). Thus, by (8.125),

$$\|\mathcal{M}_2 h\|_s^{\mathrm{Lip}(\gamma)} + \|\mathcal{M}_1^{-1} h\|_s^{\mathrm{Lip}(\gamma)} \leq_s \|h\|_{s+3}^{\mathrm{Lip}(\gamma)} + \varepsilon\gamma^{-1} \|\mathfrak{I}_\delta\|_{s+\sigma+3}^{\mathrm{Lip}(\gamma)} \|h\|_{s_0}^{\mathrm{Lip}(\gamma)},$$

and (8.124) follows, using also (6.9). □

9. The Nash–Moser nonlinear iteration

In this section we prove Theorem 5.1. It will be a consequence of the Nash–Moser Theorem 9.1 below.

Consider the finite-dimensional subspaces

$$E_n := \big\{ \mathfrak{I}(\varphi) = (\Theta, y, z)(\varphi) : \Theta = \Pi_n \Theta, \ y = \Pi_n y, \ z = \Pi_n z \big\}$$

where $N_n := N_0^{\chi^n}$ are introduced in (8.114), and Π_n are the projectors (which, with a small abuse of notation, we denote with the same symbol)

$$\begin{aligned} \Pi_n \Theta(\varphi) &:= \sum_{|l|<N_n} \Theta_l e^{\mathrm{i} l \cdot \varphi}, \quad \Pi_n y(\varphi) := \sum_{|l|<N_n} y_l e^{\mathrm{i} l \cdot \varphi}, \quad \text{where } \Theta(\varphi) = \sum_{l \in \mathbb{Z}^\nu} \Theta_l e^{\mathrm{i} l \cdot \varphi}, \quad y(\varphi) = \sum_{l \in \mathbb{Z}^\nu} y_l e^{\mathrm{i} l \cdot \varphi}, \\ \Pi_n z(\varphi, x) &:= \sum_{|(l,j)|<N_n} z_{lj} e^{\mathrm{i}(l \cdot \varphi + jx)}, \quad \text{where } z(\varphi, x) = \sum_{l \in \mathbb{Z}^\nu, j \in S^c} z_{lj} e^{\mathrm{i}(l \cdot \varphi + jx)}. \end{aligned} \tag{9.1}$$

We define $\Pi_n^\perp := I - \Pi_n$. The classical smoothing properties hold: for all $\alpha, s \geq 0$,

$$\|\Pi_n \mathfrak{I}\|_{s+\alpha}^{\mathrm{Lip}(\gamma)} \leq N_n^\alpha \|\mathfrak{I}\|_s^{\mathrm{Lip}(\gamma)}, \ \forall \mathfrak{I}(\omega) \in H^s, \quad \|\Pi_n^\perp \mathfrak{I}\|_s^{\mathrm{Lip}(\gamma)} \leq N_n^{-\alpha} \|\mathfrak{I}\|_{s+\alpha}^{\mathrm{Lip}(\gamma)}, \ \forall \mathfrak{I}(\omega) \in H^{s+\alpha}. \tag{9.2}$$

We define the constants

$$\mu_1 := 3\mu + 9, \qquad \alpha := 3\mu_1 + 1, \qquad \alpha_1 := (\alpha - 3\mu)/2, \tag{9.3}$$

$$\kappa := 3\big(\mu_1 + \rho^{-1}\big) + 1, \qquad \beta_1 := 6\mu_1 + 3\rho^{-1} + 3, \qquad 0 < \rho < \frac{1-3a}{C_1(1+a)}, \tag{9.4}$$

where $\mu := \mu(\tau, \nu)$ is the "loss of regularity" defined in Theorem 6.10 (see (6.41)) and C_1 is fixed below.

Theorem 9.1 *(Nash–Moser). Assume that $f \in C^q$ with $q > S := s_0 + \beta_1 + \mu + 3$. Let $\tau \geq \nu + 2$. Then there exist $C_1 > \max\{\mu_1 + \alpha, C_0\}$ (where $C_0 := C_0(\tau, \nu)$ is the one in Theorem 8.25), $\delta_0 := \delta_0(\tau, \nu) > 0$ such that, if*

$$N_0^{C_1} \varepsilon^{b_*+1} \gamma^{-2} < \delta_0, \quad \gamma := \varepsilon^{2+a} = \varepsilon^{2b}, \quad N_0 := (\varepsilon\gamma^{-1})^\rho, \quad b_* := 6 - 2b, \tag{9.5}$$

then, for all $n \geq 0$:

$(\mathcal{P}1)_n$ *there exists a function $(\mathfrak{I}_n, \zeta_n) : \mathcal{G}_n \subseteq \Omega_\varepsilon \to E_{n-1} \times \mathbb{R}^\nu$, $\omega \mapsto (\mathfrak{I}_n(\omega), \zeta_n(\omega))$, $(\mathfrak{I}_0, \zeta_0) := 0$, $E_{-1} := \{0\}$, satisfying $|\zeta_n|^{\mathrm{Lip}(\gamma)} \leq C\|\mathcal{F}(U_n)\|_{s_0}^{\mathrm{Lip}(\gamma)}$,*

$$\|\mathfrak{I}_n\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1}, \quad \|\mathcal{F}(U_n)\|_{s_0+\mu+3}^{\mathrm{Lip}(\gamma)} \leq C_* \varepsilon^{b_*}, \tag{9.6}$$

where $U_n := (i_n, \zeta_n)$ with $i_n(\varphi) = (\varphi, 0, 0) + \mathfrak{I}_n(\varphi)$. The sets $\mathcal{G}_n$ are defined inductively by:

$$\mathcal{G}_0 := \big\{\omega \in \Omega_\varepsilon : |\omega \cdot l| \geq 2\gamma \langle l \rangle^{-\tau}, \ \forall l \in \mathbb{Z}^\nu \setminus \{0\}\big\},$$

$$\mathcal{G}_{n+1} := \left\{\omega \in \mathcal{G}_n : |\mathrm{i}\omega \cdot l + \mu_j^\infty(i_n) - \mu_k^\infty(i_n)| \geq \frac{2\gamma_n |j^3 - k^3|}{\langle l \rangle^\tau}, \ \forall j, k \in S^c \cup \{0\}, \ l \in \mathbb{Z}^\nu\right\}, \tag{9.7}$$

where $\gamma_n := \gamma(1 + 2^{-n})$ and $\mu_j^\infty(\omega) := \mu_j^\infty(\omega, i_n(\omega))$ are defined in (8.118) (and $\mu_0^\infty(\omega) = 0$).
The differences $\widehat{\mathfrak{I}}_n := \mathfrak{I}_n - \mathfrak{I}_{n-1}$ (where we set $\widehat{\mathfrak{I}}_0 := 0$) is defined on $\mathcal{G}_n$, and satisfy

$$\|\widehat{\mathfrak{I}}_1\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1}, \quad \|\widehat{\mathfrak{I}}_n\|_{s_0+\mu}^{\mathrm{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1} N_{n-1}^{-\alpha_1}, \quad \forall n > 1. \tag{9.8}$$

$(\mathcal{P}2)_n$ $\|\mathcal{F}(U_n)\|_{s_0}^{\mathrm{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} N_{n-1}^{-\alpha}$ *where we set $N_{-1} := 1$.*
$(\mathcal{P}3)_n$ (High norms). $\|\mathfrak{I}_n\|_{s_0+\beta_1}^{\mathrm{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} \gamma^{-1} N_{n-1}^\kappa$ *and* $\|\mathcal{F}(U_n)\|_{s_0+\beta_1}^{\mathrm{Lip}(\gamma)} \leq C_* \varepsilon^{b_*} N_{n-1}^\kappa$.
$(\mathcal{P}4)_n$ (Measure). *The measure of the "Cantor-like" sets $\mathcal{G}_n$ satisfies*

$$|\Omega_\varepsilon \setminus \mathcal{G}_0| \leq C_* \varepsilon^{2(\nu-1)} \gamma, \quad \big|\mathcal{G}_n \setminus \mathcal{G}_{n+1}\big| \leq C_* \varepsilon^{2(\nu-1)} \gamma N_{n-1}^{-1}. \tag{9.9}$$

All the Lip norms are defined on $\mathcal{G}_n$, namely $\| \ \|_s^{\mathrm{Lip}(\gamma)} = \| \ \|_{s,\mathcal{G}_n}^{\mathrm{Lip}(\gamma)}$.

Proof. To simplify notations, in this proof we denote $\| \ \|^{\mathrm{Lip}(\gamma)}$ by $\| \ \|$. We first prove $(\mathcal{P}1, 2, 3)_n$.

STEP 1: *Proof of* $(\mathcal{P}1, 2, 3)_0$. Recalling (5.6) we have $\|\mathcal{F}(U_0)\|_s = \|\mathcal{F}(\varphi, 0, 0, 0)\|_s = \|X_P(\varphi, 0, 0)\|_s \leq_s \varepsilon^{6-2b}$ by (5.15). Hence (recall that $b_* = 6 - 2b$) the smallness conditions in $(\mathcal{P}1)_0$–$(\mathcal{P}3)_0$ hold taking $C_* := C_*(s_0 + \beta_1)$ large enough.

STEP 2: *Assume that $(\mathcal{P}1, 2, 3)_n$ hold for some $n \geq 0$, and prove $(\mathcal{P}1, 2, 3)_{n+1}$.* By (9.5) and (9.4),

$$N_0^{C_1} \varepsilon^{b_*+1} \gamma^{-2} = N_0^{C_1} \varepsilon^{1-3a} = \varepsilon^{1-3a-\rho C_1(1+a)} < \delta_0$$

for ε small enough, and the smallness condition (8.117) holds. Moreover (9.6) imply (6.4) (and so (7.8)) and Theorem 8.27 applies. Hence the operator $\mathcal{L}_\omega := \mathcal{L}_\omega(\omega, i_n(\omega))$ defined in (6.32) is invertible for all $\omega \in \mathcal{G}_{n+1}$ and the last estimate in (8.124) holds. This means that the assumption (6.33) of Theorem 6.10 is verified with

$\Omega_\infty = \mathcal{G}_{n+1}$. By Theorem 6.10 there exists an approximate inverse $\mathbf{T}_n(\omega) := \mathbf{T}_0(\omega, i_n(\omega))$ of the linearized operator $L_n(\omega) := d_{i,\zeta}\mathcal{F}(\omega, i_n(\omega))$, satisfying (6.41). Thus, using also (9.5), (9.6),

$$\|\mathbf{T}_n g\|_s \leq_s \gamma^{-1}\big(\|g\|_{s+\mu} + \varepsilon\gamma^{-1}\|\mathfrak{I}_n\|_{s+\mu}\|g\|_{s_0+\mu}\big) \tag{9.10}$$

$$\|\mathbf{T}_n g\|_{s_0} \leq_{s_0} \gamma^{-1}\|g\|_{s_0+\mu} \tag{9.11}$$

and, by (6.42), using also (9.6), (9.5), (9.2),

$$\begin{aligned}\|\big(L_n \circ \mathbf{T}_n - I\big)g\|_s \leq_s \varepsilon^{2b-1}\gamma^{-2}\big(&\|\mathcal{F}(U_n)\|_{s_0+\mu}\|g\|_{s+\mu} + \|\mathcal{F}(U_n)\|_{s+\mu}\|g\|_{s_0+\mu} \\ &+ \varepsilon\gamma^{-1}\|\mathfrak{I}_n\|_{s+\mu}\|\mathcal{F}(U_n)\|_{s_0+\mu}\|g\|_{s_0+\mu}\big),\end{aligned} \tag{9.12}$$

$$\begin{aligned}\|\big(L_n \circ \mathbf{T}_n - I\big)g\|_{s_0} &\leq_{s_0} \varepsilon^{2b-1}\gamma^{-2}\|\mathcal{F}(U_n)\|_{s_0+\mu}\|g\|_{s_0+\mu} \\ &\leq_{s_0} \varepsilon^{2b-1}\gamma^{-2}\big(\|\Pi_n\mathcal{F}(U_n)\|_{s_0+\mu} + \|\Pi_n^\perp\mathcal{F}(U_n)\|_{s_0+\mu}\big)\|g\|_{s_0+\mu} \\ &\leq_{s_0} \varepsilon^{2b-1}\gamma^{-2}N_n^{\mu}\big(\|\mathcal{F}(U_n)\|_{s_0} + N_n^{-\beta_1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}\big)\|g\|_{s_0+\mu}.\end{aligned} \tag{9.13}$$

Then, for all $\omega \in \mathcal{G}_{n+1}$, $n \geq 0$, we define

$$U_{n+1} := U_n + H_{n+1}, \quad H_{n+1} := (\widehat{\mathfrak{I}}_{n+1}, \widehat{\zeta}_{n+1}) := -\widetilde{\Pi}_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) \in E_n \times \mathbb{R}^\nu, \tag{9.14}$$

where $\widetilde{\Pi}_n(\mathfrak{I}, \zeta) := (\Pi_n\mathfrak{I}, \zeta)$ with Π_n in (9.1). Since $L_n := d_{i,\zeta}\mathcal{F}(i_n)$, we write $\mathcal{F}(U_{n+1}) = \mathcal{F}(U_n) + L_nH_{n+1} + Q_n$, where

$$Q_n := Q(U_n, H_{n+1}), \quad Q(U_n, H) := \mathcal{F}(U_n + H) - \mathcal{F}(U_n) - L_nH, \quad H \in E_n \times \mathbb{R}^\nu. \tag{9.15}$$

Then, by the definition of H_{n+1} in (9.14), and writing $\widetilde{\Pi}_n^\perp(\mathfrak{I}, \zeta) := (\Pi_n^\perp\mathfrak{I}, 0)$, we have

$$\begin{aligned}\mathcal{F}(U_{n+1}) &= \mathcal{F}(U_n) - L_n\widetilde{\Pi}_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + Q_n = \mathcal{F}(U_n) - L_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + L_n\widetilde{\Pi}_n^\perp\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + Q_n \\ &= \mathcal{F}(U_n) - \Pi_nL_n\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + (L_n\widetilde{\Pi}_n^\perp - \Pi_n^\perp L_n)\mathbf{T}_n\Pi_n\mathcal{F}(U_n) + Q_n \\ &= \Pi_n^\perp\mathcal{F}(U_n) + R_n + Q_n + Q_n'\end{aligned} \tag{9.16}$$

where

$$R_n := (L_n\widetilde{\Pi}_n^\perp - \Pi_n^\perp L_n)\mathbf{T}_n\Pi_n\mathcal{F}(U_n), \qquad Q_n' := -\Pi_n(L_n\mathbf{T}_n - I)\Pi_n\mathcal{F}(U_n). \tag{9.17}$$

Lemma 9.2. *Define*

$$w_n := \varepsilon\gamma^{-2}\|\mathcal{F}(U_n)\|_{s_0}, \quad B_n := \varepsilon\gamma^{-1}\|\mathfrak{I}_n\|_{s_0+\beta_1} + \varepsilon\gamma^{-2}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}. \tag{9.18}$$

Then there exists $K := K(s_0, \beta_1) > 0$ such that, for all $n \geq 0$, setting $\mu_1 := 3\mu + 9$ (see (9.3)),

$$w_{n+1} \leq KN_n^{\mu_1 + \frac{1}{\rho} - \beta_1}B_n + KN_n^{\mu_1}w_n^2, \qquad B_{n+1} \leq KN_n^{\mu_1 + \frac{1}{\rho}}B_n. \tag{9.19}$$

Proof. We estimate separately the terms Q_n in (9.15) and Q_n', R_n in (9.17).
Estimate of Q_n. By (9.15), (5.6), (5.20) and (9.6), (9.2), we have the quadratic estimates

$$\|Q(U_n, H)\|_s \leq_s \varepsilon\big(\|\widehat{\mathfrak{I}}\|_{s+3}\|\widehat{\mathfrak{I}}\|_{s_0+3} + \|\mathfrak{I}_n\|_{s+3}\|\widehat{\mathfrak{I}}\|_{s_0+3}^2\big) \tag{9.20}$$

$$\|Q(U_n, H)\|_{s_0} \leq_{s_0} \varepsilon N_n^6\|\widehat{\mathfrak{I}}\|_{s_0}^2, \quad \forall\widehat{\mathfrak{I}} \in E_n. \tag{9.21}$$

Now by the definition of H_{n+1} in (9.14) and (9.2), (9.10), (9.11), (9.6), we get

$$\begin{aligned}\|\widehat{\mathfrak{I}}_{n+1}\|_{s_0+\beta_1} &\leq_{s_0+\beta_1} N_n^{\mu}\big(\gamma^{-1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1} + \varepsilon\gamma^{-2}\|\mathcal{F}(U_n)\|_{s_0+\mu}\{\|\mathfrak{I}_n\|_{s_0+\beta_1} + \gamma^{-1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}\}\big) \\ &\leq_{s_0+\beta} N_n^{\mu}\big(\gamma^{-1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1} + \|\mathfrak{I}_n\|_{s_0+\beta_1}\big),\end{aligned} \tag{9.22}$$

$$\|\widehat{\mathfrak{I}}_{n+1}\|_{s_0} \leq_{s_0} \gamma^{-1}N_n^{\mu}\|\mathcal{F}(U_n)\|_{s_0}. \tag{9.23}$$

Then the term Q_n in (9.15) satisfies, by (9.20), (9.21), (9.22), (9.23), (9.5), (9.6), $(\mathcal{P}2)_n$, (9.3),

$$\|Q_n\|_{s_0+\beta_1} \le_{s_0+\beta_1} N_n^{2\mu+9}\gamma\big(\gamma^{-1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}+\|\mathfrak{I}_n\|_{s_0+\beta_1}\big)\,, \tag{9.24}$$

$$\|Q_n\|_{s_0} \le_{s_0} N_n^{2\mu+6}\varepsilon\gamma^{-2}\|\mathcal{F}(U_n)\|_{s_0}^2\,. \tag{9.25}$$

Estimate of Q_n'. The bounds (9.12), (9.13), (9.2), (9.3), (9.6) imply

$$\|Q_n'\|_{s_0+\beta_1} \le_{s_0+\beta_1} \varepsilon^5\gamma^{-2}N_n^{2\mu}\big(\|\mathcal{F}(U_n)\|_{s_0+\beta_1}+\varepsilon\gamma^{-1}\|\mathfrak{I}_n\|_{s_0+\beta_1}\|\mathcal{F}(U_n)\|_{s_0}\big)\,, \tag{9.26}$$

$$\|Q_n'\|_{s_0} \le_{s_0} \varepsilon^{2b-1}\gamma^{-2}N_n^{2\mu}\big(\|\mathcal{F}(U_n)\|_{s_0}+N_n^{-\beta_1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}\big)\|\mathcal{F}(U_n)\|_{s_0}\,. \tag{9.27}$$

Estimate of R_n. For $H := (\widehat{\mathfrak{I}}, \widehat{\zeta})$ we have $(L_n\widetilde{\Pi}_n^{\perp} - \Pi_n^{\perp}L_n)H = [\bar D_n, \Pi_n^{\perp}]\widehat{\mathfrak{I}} = [\Pi_n, \bar D_n]\widehat{\mathfrak{I}}$ where $\bar D_n := d_i X_{H_\varepsilon}(i_n) + (0,0,\partial_{xxx})$. Thus Lemma 5.3, (9.6), (9.2) and (5.19) imply

$$\|(L_n\widetilde{\Pi}_n^{\perp} - \Pi_n^{\perp}L_n)H\|_{s_0} \le_{s_0+\beta_1} \varepsilon N_n^{-\beta_1+\mu+3}\big(\|\widehat{\mathfrak{I}}\|_{s_0+\beta_1-\mu}+\|\mathfrak{I}_n\|_{s_0+\beta_1-\mu}\|\widehat{\mathfrak{I}}\|_{s_0+3}\big)\,, \tag{9.28}$$

$$\|(L_n\widetilde{\Pi}_n^{\perp} - \Pi_n^{\perp}L_n)H\|_{s_0+\beta_1} \le_{s} \varepsilon N_n^{\mu+3}\big(\|\widehat{\mathfrak{I}}\|_{s_0+\beta_1-\mu}+\|\mathfrak{I}_n\|_{s_0+\beta_1-\mu}\|\widehat{\mathfrak{I}}\|_{s_0+3}\big)\,. \tag{9.29}$$

Hence, applying (9.10), (9.28), (9.29), (9.5), (9.6), (9.2), the term R_n defined in (9.17) satisfies

$$\|R_n\|_{s_0} \le_{s_0+\beta_1} N_n^{\mu+6-\beta_1}(\varepsilon\gamma^{-1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}+\varepsilon\|\mathfrak{I}_n\|_{s_0+\beta_1})\,, \tag{9.30}$$

$$\|R_n\|_{s_0+\beta_1} \le_{s_0+\beta_1} N_n^{\mu+6}(\varepsilon\gamma^{-1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}+\varepsilon\|\mathfrak{I}_n\|_{s_0+\beta_1})\,. \tag{9.31}$$

Estimate of $\mathcal{F}(U_{n+1})$. By (9.16) and (9.24), (9.25), (9.26), (9.27), (9.30), (9.31), (9.5), (9.6), we get

$$\|\mathcal{F}(U_{n+1})\|_{s_0} \le_{s_0+\beta_1} N_n^{\mu_1-\beta_1}(\varepsilon\gamma^{-1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}+\varepsilon\|\mathfrak{I}_n\|_{s_0+\beta_1}) + N_n^{\mu_1}\varepsilon\gamma^{-2}\|\mathcal{F}(U_n)\|_{s_0}^2\,, \tag{9.32}$$

$$\|\mathcal{F}(U_{n+1})\|_{s_0+\beta_1} \le_{s_0+\beta_1} N_n^{\mu_1}(\varepsilon\gamma^{-1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1}+\varepsilon\|\mathfrak{I}_n\|_{s_0+\beta_1})\,, \tag{9.33}$$

where $\mu_1 := 3\mu + 9$.
Estimate of $\mathfrak{I}_{n+1}$. Using (9.22) the term $\mathfrak{I}_{n+1} = \mathfrak{I}_n + \widehat{\mathfrak{I}}_{n+1}$ is bounded by

$$\|\mathfrak{I}_{n+1}\|_{s_0+\beta_1} \le_{s_0+\beta_1} N_n^{\mu}(\|\mathfrak{I}_n\|_{s_0+\beta_1}+\gamma^{-1}\|\mathcal{F}(U_n)\|_{s_0+\beta_1})\,. \tag{9.34}$$

Finally, recalling (9.18), the inequalities (9.19) follow by (9.32)–(9.34), (9.6) and $\varepsilon\gamma^{-1} = N_0^{1/\rho} \le N_n^{1/\rho}$. □

Proof of $(\mathcal{P}3)_{n+1}$. By (9.19) and $(\mathcal{P}3)_n$,

$$B_{n+1} \le KN_n^{\mu_1+\frac{1}{\rho}}B_n \le 2C_*K\varepsilon^{b_*+1}\gamma^{-2}N_n^{\mu_1+\frac{1}{\rho}}N_{n-1}^{\kappa} \le C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{\kappa}\,, \tag{9.35}$$

provided $2KN_n^{\mu_1+\frac{1}{\rho}-\kappa}N_{n-1}^{\kappa} \le 1$, $\forall n \ge 0$. This inequality holds by (9.4), taking N_0 large enough (i.e. ε small enough). By (9.18), the bound $B_{n+1} \le C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{\kappa}$ implies $(\mathcal{P}3)_{n+1}$.

Proof of $(\mathcal{P}2)_{n+1}$. Using (9.19), (9.18) and $(\mathcal{P}2)_n$, $(\mathcal{P}3)_n$, we get

$$w_{n+1} \le KN_n^{\mu_1+\frac{1}{\rho}-\beta_1}B_n + KN_n^{\mu_1}w_n^2 \le KN_n^{\mu_1+\frac{1}{\rho}-\beta_1}2C_*\varepsilon^{b_*+1}\gamma^{-2}N_{n-1}^{\kappa} + KN_n^{\mu_1}(C_*\varepsilon^{b_*+1}\gamma^{-2}N_{n-1}^{-\alpha})^2$$

which is $\le C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{-\alpha}$ provided that

$$4KN_n^{\mu_1+\frac{1}{\rho}-\beta_1+\alpha}N_{n-1}^{\kappa} \le 1, \quad 2KC_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{\mu_1+\alpha}N_{n-1}^{-2\alpha} \le 1\,, \quad \forall n \ge 0. \tag{9.36}$$

The inequalities in (9.36) hold by (9.3)–(9.4), (9.5), $C_1 > \mu_1 + \alpha$, taking δ_0 in (9.5) small enough. By (9.18), the inequality $w_{n+1} \le C_*\varepsilon^{b_*+1}\gamma^{-2}N_n^{-\alpha}$ implies $(\mathcal{P}2)_{n+1}$.

Proof of $(\mathcal{P}1)_{n+1}$. The bound (9.8) for $\widehat{\mathfrak{I}}_1$ follows by (9.14), (9.10) (for $s = s_0 + \mu$) and $\|\mathcal{F}(U_0)\|_{s_0+2\mu} = \|\mathcal{F}(\varphi,0,0,0)\|_{s_0+2\mu} \le_{s_0+2\mu} \varepsilon^{b_*}$. The bound (9.8) for $\widehat{\mathfrak{I}}_{n+1}$ follows by (9.2), (9.23), $(\mathcal{P}2)_n$, (9.3). It remains to prove that (9.6) holds at the step $n+1$. We have

$$\|\mathfrak{I}_{n+1}\|_{s_0+\mu} \le \sum_{k=1}^{n+1}\|\widehat{\mathfrak{I}}_k\|_{s_0+\mu} \le C_*\varepsilon^{b_*}\gamma^{-1}\sum_{k\ge1}N_{k-1}^{-\alpha_1} \le C_*\varepsilon^{b_*}\gamma^{-1} \tag{9.37}$$

for N_0 large enough, i.e. ε small. Moreover, using (9.2), $(\mathcal{P}2)_{n+1}$, $(\mathcal{P}3)_{n+1}$, (9.3), we get

$$
\begin{aligned}
\|\mathcal{F}(U_{n+1})\|_{s_0+\mu+3} &\le N_n^{\mu+3}\|\mathcal{F}(U_{n+1})\|_{s_0} + N_n^{\mu+3-\beta_1}\|\mathcal{F}(U_{n+1})\|_{s_0+\beta_1} \\
&\le C_*\varepsilon^{b_*}N_n^{\mu+3-\alpha} + C_*\varepsilon^{b_*}N_n^{\mu+3-\beta_1+\kappa} \le C_*\varepsilon^{b_*},
\end{aligned}
$$

which is the second inequality in (9.6) at the step $n+1$. The bound $|\zeta_{n+1}|^{\mathrm{Lip}(\gamma)} \le C\|\mathcal{F}(U_{n+1})\|_{s_0}^{\mathrm{Lip}(\gamma)}$ is a consequence of Lemma 6.1 (it is not inductive).

STEP 3: *Prove* $(\mathcal{P}4)_n$ *for all* $n \ge 0$. For all $n \ge 0$,

$$
\mathcal{G}_n \setminus \mathcal{G}_{n+1} = \bigcup_{l\in\mathbb{Z}^\nu,\ j,k\in S^c\cup\{0\}} R_{ljk}(i_n) \tag{9.38}
$$

where

$$
R_{ljk}(i_n) := \left\{\omega \in \mathcal{G}_n \,:\, |\mathrm{i}\omega\cdot l + \mu_j^\infty(i_n) - \mu_k^\infty(i_n)| < 2\gamma_n|j^3-k^3|\,\langle l\rangle^{-\tau}\right\}. \tag{9.39}
$$

Notice that $R_{ljk}(i_n) = \emptyset$ if $j=k$, so that we suppose in the sequel that $j \ne k$.

Lemma 9.3. *For all* $n \ge 1$, $|l| \le N_{n-1}$, *the set* $R_{ljk}(i_n) \subseteq R_{ljk}(i_{n-1})$.

Proof. Like Lemma 5.2 in [2] (with ω in the role of $\lambda\bar\omega$, and N_{n-1} instead of N_n). □

By definition, $R_{ljk}(i_n) \subseteq \mathcal{G}_n$ (see (9.39)) and Lemma 9.3 implies that, for all $n \ge 1$, $|l| \le N_{n-1}$, the set $R_{ljk}(i_n) \subseteq R_{ljk}(i_{n-1})$. On the other hand $R_{ljk}(i_{n-1}) \cap \mathcal{G}_n = \emptyset$ (see (9.7)). As a consequence, for all $|l| \le N_{n-1}$, $R_{ljk}(i_n) = \emptyset$ and, by (9.38),

$$
\mathcal{G}_n \setminus \mathcal{G}_{n+1} \subseteq \bigcup_{|l|>N_{n-1},\ j,k\in S^c\cup\{0\}} R_{ljk}(i_n) \qquad \forall n \ge 1. \tag{9.40}
$$

Lemma 9.4. *Let* $n \ge 0$. *If* $R_{ljk}(i_n) \ne \emptyset$ *then* $|l| \ge C|j^3-k^3| \ge \frac{1}{2}C(j^2+k^2)$ *for some* $C > 0$.

Proof. Like Lemma 5.3 in [2]. The only difference is that ω is not constrained to a fixed direction. Note also that $|j^3-k^3| \ge (j^2+k^2)/2$, $\forall j \ne k$. □

By usual arguments (e.g. see Lemma 5.4 in [2]), using Lemma 9.4 and (8.119) we have:

Lemma 9.5. *For all* $n \ge 0$, *the measure* $|R_{ljk}(i_n)| \le C\varepsilon^{2(\nu-1)}\gamma\langle l\rangle^{-\tau}$.

By (9.38) and Lemmata 9.4, 9.5 we get

$$
|\mathcal{G}_0 \setminus \mathcal{G}_1| \le \sum_{l\in\mathbb{Z}^\nu,|j|,|k|\le C|l|^{1/2}} |R_{ljk}(i_0)| \le \sum_{l\in\mathbb{Z}^\nu} \frac{C\varepsilon^{2(\nu-1)}\gamma}{\langle l\rangle^{\tau-1}} \le C'\varepsilon^{2(\nu-1)}\gamma\,.
$$

For $n \ge 1$, by (9.40),

$$
|\mathcal{G}_n \setminus \mathcal{G}_{n+1}| \le \sum_{|l|>N_{n-1},|j|,|k|\le C|l|^{1/2}} |R_{ljk}(i_n)| \le \sum_{|l|>N_{n-1}} \frac{C\varepsilon^{2(\nu-1)}\gamma}{\langle l\rangle^{\tau-1}} \le C'\varepsilon^{2(\nu-1)}\gamma N_{n-1}^{-1}
$$

because $\tau \ge \nu+2$. The estimate $|\Omega_\varepsilon \setminus \mathcal{G}_0| \le C\varepsilon^{2(\nu-1)}\gamma$ is elementary. Thus (9.9) is proved. □

Proof of Theorem 5.1 concluded. Theorem 9.1 implies that the sequence $(\mathfrak{I}_n, \zeta_n)$ is well defined for $\omega \in \mathcal{G}_\infty := \cap_{n\ge0}\mathcal{G}_n$, that $\mathfrak{I}_n$ is a Cauchy sequence in $\|\ \|_{s_0+\mu,\mathcal{G}_\infty}^{\mathrm{Lip}(\gamma)}$, see (9.8), and $|\zeta_n|^{\mathrm{Lip}(\gamma)} \to 0$. Therefore $\mathfrak{I}_n$ converges to a limit $\mathfrak{I}_\infty$ in norm $\|\ \|_{s_0+\mu,\mathcal{G}_\infty}^{\mathrm{Lip}(\gamma)}$ and, by $(\mathcal{P}2)_n$, for all $\omega \in \mathcal{G}_\infty$, $i_\infty(\varphi) := (\varphi,0,0) + \mathfrak{I}_\infty(\varphi)$, is a solution of

$$
\mathcal{F}(i_\infty, 0) = 0 \quad \text{with} \quad \|\mathfrak{I}_\infty\|_{s_0+\mu,\mathcal{G}_\infty}^{\mathrm{Lip}(\gamma)} \le C\varepsilon^{6-2b}\gamma^{-1}
$$

by (9.6) (recall that $b_* := 6 - 2b$). Therefore $\varphi \mapsto i_\infty(\varphi)$ is an invariant torus for the Hamiltonian vector field X_{H_ε} (see (5.5)). By (9.9),

$$|\Omega_\varepsilon \setminus \mathcal{G}_\infty| \le |\Omega_\varepsilon \setminus \mathcal{G}_0| + \sum_{n\ge 0} |\mathcal{G}_n \setminus \mathcal{G}_{n+1}| \le 2C_* \varepsilon^{2(\nu-1)}\gamma + C_*\varepsilon^{2(\nu-1)}\gamma \sum_{n\ge 1} N_{n-1}^{-1} \le C\varepsilon^{2(\nu-1)}\gamma\,.$$

The set Ω_ε in (5.2) has measure $|\Omega_\varepsilon| = O(\varepsilon^{2\nu})$. Hence $|\Omega_\varepsilon \setminus \mathcal{G}_\infty|/|\Omega_\varepsilon| \to 0$ as $\varepsilon \to 0$ because $\gamma = o(\varepsilon^2)$, and therefore the measure of $\mathcal{C}_\varepsilon := \mathcal{G}_\infty$ satisfies (5.11).

In order to complete the proof of Theorem 5.1 we show the linear stability of the solution $i_\infty(\omega t)$. By Section 6 the system obtained linearizing the Hamiltonian vector field X_{H_ε} at a quasi-periodic solution $i_\infty(\omega t)$ is conjugated to the linear Hamiltonian system

$$\begin{cases} \dot\psi &= K_{20}(\omega t)\eta + K_{11}^T(\omega t) w \\ \dot\eta &= 0 \\ \dot w - \partial_x K_{02}(\omega t) w &= \partial_x K_{11}(\omega t)\eta \end{cases} \tag{9.41}$$

(recall that the torus i_∞ is isotropic and the transformed nonlinear Hamiltonian system is (6.21) where $K_{00}, K_{10}, K_{01} = 0$, see Remark 6.5). In Section 8 we have proved the reducibility of the linear system $\dot w - \partial_x K_{02}(\omega t) w$, conjugating the last equation in (9.41) to a diagonal system

$$\dot v_j + \mu_j^\infty v_j = f_j(\omega t)\,, \quad j \in S^c\,, \quad \mu_j^\infty \in \mathrm{i}\mathbb{R}\,, \tag{9.42}$$

see (8.121), and $f(\varphi, x) = \sum_{j\in S^c} f_j(\varphi) e^{\mathrm{i}jx} \in H^s_{S^\perp}(\mathbb{T}^{\nu+1})$. Thus (9.41) is stable. Indeed the actions $\eta(t) = \eta_0 \in \mathbb{R}$, $\forall t \in \mathbb{R}$. Moreover the solutions of the non-homogeneous equation (9.42) are

$$v_j(t) = c_j e^{\mu_j^\infty t} + \tilde v_j(t)\,, \quad \text{where} \quad \tilde v_j(t) := \sum_{l\in\mathbb{Z}^\nu} \frac{f_{jl}\, e^{\mathrm{i}\omega\cdot lt}}{\mathrm{i}\omega\cdot l + \mu_j^\infty}$$

is a quasi-periodic solution (recall that the first Melnikov conditions (8.123) hold at a solution). As a consequence (recall also $\mu_j^\infty \in \mathrm{i}\mathbb{R}$) the Sobolev norm of the solution of (9.42) with initial condition $v(0) = \sum_{j\in S^c} v_j(0)e^{\mathrm{i}jx} \in H^{s_0}(\mathbb{T}_x)$, $s_0 < s$, does not increase in time. □

Construction of the set S of tangential sites. We finally prove that, for any $\nu \ge 1$, the set S in (1.8) satisfying ($\mathtt{S}1$)–($\mathtt{S}2$) can be constructed inductively with only a *finite* number of restriction at any step of the induction.

First, fix any integer $\bar\jmath_1 \ge 1$. Then the set $J_1 := \{\pm\bar\jmath_1\}$ trivially satisfies ($\mathtt{S}1$)–($\mathtt{S}2$). Then, assume that we have fixed n distinct positive integers $\bar\jmath_1, \dots, \bar\jmath_n$, $n \ge 1$, such that the set $J_n := \{\pm\bar\jmath_1, \dots, \pm\bar\jmath_n\}$ satisfies ($\mathtt{S}1$)–($\mathtt{S}2$). We describe how to choose another positive integer $\bar\jmath_{n+1}$, which is different from all $j \in J_n$, such that $J_{n+1} := J_n \cup \{\pm\bar\jmath_{n+1}\}$ also satisfies ($\mathtt{S}1$), ($\mathtt{S}2$).

Let us begin with analyzing ($\mathtt{S}1$). A set of 3 elements $j_1, j_2, j_3 \in J_{n+1}$ can be of these types: (i) all "old" elements $j_1, j_2, j_3 \in J_n$; (ii) two "old" elements $j_1, j_2 \in J_n$ and one "new" element $j_3 = \sigma_3\bar\jmath_{n+1}$, $\sigma_3 = \pm 1$; (iii) one "old" element $j_1 \in J_n$ and two "new" elements $j_2 = \sigma_2\bar\jmath_{n+1}$, $j_3 = \sigma_3\bar\jmath_{n+1}$, with $\sigma_2, \sigma_3 = \pm 1$; (iv) all "new" elements $j_i = \sigma_i\bar\jmath_{n+1}$, $\sigma_i = \pm 1$, $i = 1, 2, 3$.

In case (i), the sum $j_1 + j_2 + j_3$ is nonzero by inductive assumption. In case (ii), $j_1 + j_2 + j_3$ is nonzero provided $\bar\jmath_{n+1} \notin \{j_1 + j_2 : j_1, j_2 \in J_n\}$, which is a finite set. In case (iii), for $\sigma_2 + \sigma_3 = 0$ the sum $j_1 + j_2 + j_3 = j_1$ is trivially nonzero because $0 \notin J_n$, while, for $\sigma_2 + \sigma_3 \ne 0$, the sum $j_1 + j_2 + j_3 = j_1 + (\sigma_2 + \sigma_3)\bar\jmath_{n+1} \ne 0$ if $\bar\jmath_{n+1} \notin \{\frac{1}{2} j : j \in J_n\}$, which is a finite set. In case (iv), the sum $j_1 + j_2 + j_3 = (\sigma_1 + \sigma_2 + \sigma_3)\bar\jmath_{n+1} \ne 0$ because $\bar\jmath_{n+1} \ge 1$ and $\sigma_1 + \sigma_2 + \sigma_3 \in \{\pm 1, \pm 3\}$.

Now we study ($\mathtt{S}2$) for the set J_{n+1}. Denote, in short, $b := j_1^3 + j_2^3 + j_3^3 + j_4^3 - (j_1 + j_2 + j_3 + j_4)^3$.

A set of 4 elements $j_1, j_2, j_3, j_4 \in J_{n+1}$ can be of 5 types: (i) all "old" elements $j_1, j_2, j_3, j_4 \in J_n$; (ii) three "old" elements $j_1, j_2, j_3 \in J_n$ and one "new" element $j_4 = \sigma_4\bar\jmath_{n+1}$, $\sigma_4 = \pm 1$; (iii) two "old" element $j_1, j_2 \in J_n$ and two "new" elements $j_3 = \sigma_3\bar\jmath_{n+1}$, $j_4 = \sigma_4\bar\jmath_{n+1}$, with $\sigma_3, \sigma_4 = \pm 1$; (iv) one "old" element $j_1 \in J_n$ and three "new" elements $j_i = \sigma_i\bar\jmath_{n+1}$, $\sigma_i = \pm 1$, $i = 2, 3, 4$; (v) all "new" elements $j_i = \sigma_i\bar\jmath_{n+1}$, $\sigma_i = \pm 1$, $i = 1, 2, 3, 4$.

In case (i), $b \ne 0$ by inductive assumption.

In case (ii), assume that $j_1 + j_2 + j_3 + j_4 \ne 0$, and calculate

$$b = -3(j_1 + j_2 + j_3)\bar\jmath_{n+1}^2 - 3(j_1 + j_2 + j_3)^2\sigma_4\bar\jmath_{n+1} + [j_1^3 + j_2^3 + j_3^3 - (j_1 + j_2 + j_3)^3] =: p_{j_1,j_2,j_3,\sigma_4}(\bar\jmath_{n+1}).$$

This is nonzero provided $p_{j_1,j_2,j_3,\sigma_4}(\bar{\jmath}_{n+1}) \neq 0$ for all $j_1, j_2, j_3 \in J_n$, $\sigma_4 = \pm 1$. The polynomial p_{j_1,j_2,j_3,σ_4} is never identically zero because either the leading coefficient $-3(j_1 + j_2 + j_3) \neq 0$ (and, if one uses ($\mathtt{S}_3$), this is always the case), or, if $j_1 + j_2 + j_3 = 0$, then $j_1^3 + j_2^3 + j_3^3 \neq 0$ by (3.12) (using also that $0 \notin J_n$).

In case (iii), assume that $j_1 + \ldots + j_4 = j_1 + j_2 + (\sigma_3 + \sigma_4)\bar{\jmath}_{n+1} \neq 0$, and calculate

$$b = -3\alpha \bar{\jmath}_{n+1}^3 - 3\alpha^2(j_1 + j_2)\bar{\jmath}_{n+1}^2 - 3(j_1 + j_2)^2\alpha \bar{\jmath}_{n+1} - j_1 j_2(j_1 + j_2) =: q_{j_1,j_2,\alpha}(\bar{\jmath}_{n+1}),$$

where $\alpha := \sigma_3 + \sigma_4$. We impose that $q_{j_1,j_2,\alpha}(\bar{\jmath}_{n+1}) \neq 0$ for all $j_1, j_2 \in J_n$, $\alpha \in \{\pm 2, 0\}$. The polynomial $q_{j_1,j_2,\alpha}$ is never identically zero because either the leading coefficient $-3\alpha \neq 0$, or, for $\alpha = 0$, the constant term $-j_1 j_2(j_1 + j_2) \neq 0$ (recall that $0 \notin J_n$ and $j_1 + j_2 + \alpha \bar{\jmath}_{n+1} \neq 0$).

In case (iv), assume that $j_1 + \ldots + j_4 = j_1 + \alpha \bar{\jmath}_{n+1} \neq 0$, where $\alpha := \sigma_2 + \sigma_3 + \sigma_4 \in \{\pm 1, \pm 3\}$, and calculate

$$b = \alpha \bar{\jmath}_{n+1} r_{j_1,\alpha}(\bar{\jmath}_{n+1}), \quad r_{j_1,\alpha}(x) := (1 - \alpha^2)x^2 - 3\alpha j_1 x - 3j_1^2.$$

The polynomial $r_{j_1,\alpha}$ is never identically zero because $j_1 \neq 0$. We impose $r_{j_1,\alpha}(\bar{\jmath}_{n+1}) \neq 0$ for all $j_1 \in J_n$, $\alpha \in \{\pm 1, \pm 3\}$.

In case (v), assume that $j_1 + \ldots + j_4 = \alpha \bar{\jmath}_{n+1} \neq 0$, with $\alpha := \sigma_1 + \ldots + \sigma_4 \neq 0$, and calculate $b = \alpha(1 - \alpha^2)\bar{\jmath}_{n+1}^3$. This is nonzero because $\bar{\jmath}_{n+1} \geq 1$ and $\alpha \in \{\pm 2, \pm 4\}$.

We have proved that, in choosing $\bar{\jmath}_{n+1}$, there are only finitely many integers to avoid.

Conflict of interest statement

We wish to confirm that there are no conflicts of interest concerning associated with the publication "KAM for autonomous quasi-linear perturbations of KdV" by Pietro Baldi, Massimiliano Berti, Riccardo Montalto and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgements

We thank M. Procesi, P. Bolle and T. Kappeler for many useful discussions. This research was supported by the European Research Council under FP7, grant 306414, and partially by the grants STAR 2013 (UniNA and Compagnia di San Paolo) and PRIN 2012 "Variational and perturbative aspects of nonlinear differential problems".

References

[1] P. Baldi, Periodic solutions of fully nonlinear autonomous equations of Benjamin–Ono type, Ann. Inst. Henri Poincaré, Anal. Non Linéaire 30 (2013) 33–77.
[2] P. Baldi, M. Berti, R. Montalto, KAM for quasi-linear and fully nonlinear forced perturbations of Airy equation, Math. Ann. 359 (2014) 471–536.
[3] P. Baldi, M. Berti, R. Montalto, KAM for quasi-linear KdV, C. R. Acad. Sci. Paris, Ser. I 352 (2014) 603–607.
[4] M. Berti, P. Biasco, M. Procesi, KAM theory for the Hamiltonian DNLW, Ann. Sci. Éc. Norm. Supér. (4) 46 (2) (2013) 301–373.
[5] M. Berti, P. Biasco, M. Procesi, KAM theory for the reversible derivative wave equation, Arch. Ration. Mech. Anal. 212 (2014) 905–955.
[6] M. Berti, P. Bolle, Quasi-periodic solutions with Sobolev regularity of NLS on $\mathbb{T}^d$ with a multiplicative potential, Eur. J. Math. 15 (2013) 229–286.
[7] M. Berti, P. Bolle, A Nash–Moser approach to KAM theory, in: P. Guyenne, D. Nicholls, C. Sulem (Eds.), Special Volume "Hamiltonian PDEs and Applications", in: Fields Institute Communications, vol. 75, 2015.
[8] M. Berti, P. Bolle, Quasi-periodic solutions for autonomous NLW on $\mathbb{T}^d$ with a multiplicative potential, in preparation.
[9] J. Bourgain, Gibbs measures and quasi-periodic solutions for nonlinear Hamiltonian partial differential equations, in: Gelfand Math. Sem., Birkhäuser Boston, Boston, MA, 1996, pp. 23–43.
[10] J. Bourgain, Green's Function Estimates for Lattice Schrödinger Operators and Applications, Annals of Mathematics Studies, vol. 158, Princeton University Press, Princeton, 2005.
[11] W. Craig, Problèmes de petits diviseurs dans les équations aux dérivées partielles, Panoramas et Synthèses, vol. 9, Société Mathématique de France, Paris, 2000.
[12] W. Craig, C.E. Wayne, Newton's method and periodic solutions of nonlinear wave equation, Commun. Pure Appl. Math. 46 (1993) 1409–1498.
[13] L.H. Eliasson, S. Kuksin, KAM for non-linear Schrödinger equation, Ann. Math. 172 (2010) 371–435.
[14] J. Geng, X. Xu, J. You, An infinite dimensional KAM theorem and its application to the two dimensional cubic Schrödinger equation, Adv. Math. 226 (2011) 5361–5402.
[15] H. Guan, S. Kuksin, The KdV equation under periodic boundary conditions and its perturbations, Nonlinearity 27 (9) (2014) R61–R88.

[16] G. Iooss, P.I. Plotnikov, Small divisor problem in the theory of three-dimensional water gravity waves, Mem. Am. Math. Soc. 200 (940) (2009).
[17] G. Iooss, P.I. Plotnikov, Asymmetrical three-dimensional travelling gravity waves, Arch. Ration. Mech. Anal. 200 (3) (2011) 789–880.
[18] G. Iooss, P.I. Plotnikov, J.F. Toland, Standing waves on an infinitely deep perfect fluid under gravity, Arch. Ration. Mech. Anal. 177 (3) (2005) 367–478.
[19] P. Lax, Development of singularities of solutions of nonlinear hyperbolic partial differential equations, J. Math. Phys. 5 (1964) 611–613.
[20] J. Liu, X. Yuan, A KAM theorem for Hamiltonian partial differential equations with unbounded perturbations, Commun. Math. Phys. 307 (3) (2011) 629–673.
[21] T. Kappeler, J. Pöschel, KAM and KdV, Springer, 2003.
[22] S. Klainerman, A. Majda, Formation of singularities for wave equations including the nonlinear vibrating string, Commun. Pure Appl. Math. 33 (1980) 241–263.
[23] S. Kuksin, Hamiltonian perturbations of infinite-dimensional linear systems with imaginary spectrum, Funkc. Anal. Prilozh. 21 (3) (1987) 22–37, 95.
[24] S. Kuksin, A KAM theorem for equations of the Korteweg–de Vries type, Rev. Math. Phys. 10 (3) (1998) 1–64.
[25] S. Kuksin, Analysis of Hamiltonian PDEs, Oxford Lecture Series in Mathematics and Its Applications, vol. 19, Oxford University Press, 2000.
[26] S. Kuksin, J. Pöschel, Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation, Ann. Math. 2 (143) (1996) 149–179.
[27] J. Pöschel, A KAM-theorem for some nonlinear PDEs, Ann. Sc. Norm. Pisa 23 (1996) 119–148.
[28] J. Pöschel, Quasi-periodic solutions for a nonlinear wave equation, Comment. Math. Helv. 71 (2) (1996) 269–296.
[29] M. Procesi, C. Procesi, A normal form for the Schrödinger equation with analytic non-linearities, Commun. Math. Phys. 312 (2012) 501–557.
[30] C. Procesi, M. Procesi, A KAM algorithm for the completely resonant nonlinear Schrödinger equation, Adv. Math. 272 (2015) 399–470.
[31] M.E. Taylor, Pseudodifferential Operators and Nonlinear PDEs, Progress in Mathematics, Birkhäuser, 1991.
[32] W.M. Wang, Supercritical nonlinear Schrödinger equations I: quasi-periodic solutions, preprint.
[33] E. Wayne, Periodic and quasi-periodic solutions of nonlinear wave equations via KAM theory, Commun. Math. Phys. 127 (1990) 479–528.
[34] J. Zhang, M. Gao, X. Yuan, KAM tori for reversible partial differential equations, Nonlinearity 24 (2011) 1189–1228.
[35] E. Zehnder, Generalized implicit function theorems with applications to some small divisors problems I, Commun. Pure Appl. Math. 28 (1975) 91–140;
E. Zehnder, Generalized implicit function theorems with applications to some small divisors problems II, Commun. Pure Appl. Math. 29 (1976) 49–113.